

Penerapan XGBoost dan Random Forest dalam Prediksi Strategi Pemasaran Berdasarkan Data Penjualan

Dayini Nur Sabrina^{*1}, Basuki Rahmat², Eva Yulia Puspaningrum³

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur

E-mail: ^{*1}dayininursabrina@gmail.com, ²basukirahmat.if@upnjatim.ac.id,

³evapuspaningrum.if@upnjatim.ac.id

Abstrak

Penelitian ini bertujuan merancang model prediksi strategi pemasaran berbasis data transaksi penjualan dengan menerapkan algoritma XGBoost dan Random Forest secara hybrid melalui pendekatan ensemble learning. Studi dilakukan pada toko online Infinity Jar untuk memprediksi jumlah penjualan produk dan memberikan rekomendasi strategi pemasaran berupa kombinasi paket produk berdasarkan tingkat permintaan. Metode penelitian meliputi prapemrosesan data, pemodelan, tuning hyperparameter, evaluasi model menggunakan metrik Mean Squared Error (MSE), Mean Absolute Error (MAE), dan Mean Absolute Percentage Error (MAPE), serta visualisasi hasil. Model hybrid yang dibangun dengan Voting Regressor menunjukkan performa terbaik dengan MSE sebesar 5,36, MAE sebesar 1,00, dan MAPE sebesar 14,39%. Model hybrid terbukti lebih akurat dibandingkan model tunggal, yaitu Random Forest (MSE: 5,45, MAE: 1,09, MAPE: 23,48%) dan XGBoost (MSE: 6,91, MAE: 1,27, MAPE: 33,01%). Nilai metrik yang lebih rendah menunjukkan kemampuan prediksi yang lebih baik. Selain itu, penelitian ini menghasilkan 18 rekomendasi bundling produk berdasarkan klasifikasi permintaan tinggi dan rendah. Rekomendasi ini diharapkan dapat membantu optimalisasi pengelolaan stok dan meningkatkan penjualan. Dengan demikian, pendekatan hybrid terbukti efektif dalam mendukung pengambilan keputusan pemasaran berbasis data.

Kata Kunci—XGBoost, Random Forest, Prediksi Penjualan, Strategi pemasaran

Abstract

This research aims to design a marketing strategy prediction model based on sales transaction data by applying the XGBoost and Random Forest algorithms in a hybrid manner through an ensemble learning approach. The study was conducted at the Infinity Jar online store to predict the number of product sales and provide marketing strategy recommendations in the form of product package combinations based on the level of demand. Research methods include data preprocessing, modeling, hyperparameter tuning, model evaluation using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics, and visualization of results. The hybrid model built with Voting Regressor showed the best performance with MSE of 5,36, MAE of 1,00, and MAPE of 14,39%. The hybrid model proved to be more accurate than the single models, namely Random Forest (MSE: 5,45, MAE: 1,09, MAPE: 23,48%) and XGBoost (MSE: 6,91, MAE: 1,27, MAPE: 33,01%). Lower metric values indicate better prediction capabilities. In addition, this research produced 18 product bundling recommendations based on the classification of high and low demand. These recommendations are expected to help optimize stock management and increase sales. Thus, the hybrid approach proved to be effective in supporting data-driven marketing decision-making.

Keywords—XGBoost, Random Forest, Sales Prediction, Marketing Strategy

1. PENDAHULUAN

Kemajuan teknologi informasi yang terus berkembang seiring dengan meningkatnya peradaban manusia turut memberikan dampak signifikan terhadap dunia bisnis. Untuk tetap eksis dan mampu bersaing dengan para kompetitor, pelaku usaha dituntut untuk mengoptimalkan seluruh potensi yang dimilikinya [1]. Dalam kondisi persaingan yang semakin ketat, fokus usaha tidak hanya terbatas pada produk yang ditawarkan, tetapi juga perlu diarahkan pada upaya untuk memprediksi penjualan. Kemampuan dalam menganalisis data penjualan menjadi faktor kunci dalam merancang strategi pemasaran yang lebih efektif guna mempertahankan posisi di pasar yang kompetitif [2].

Dengan meningkatnya pertumbuhan *e-commerce*, semakin banyak pelaku usaha yang mulai memanfaatkan teknologi digital guna meningkatkan efisiensi dalam penjualan dan strategi pemasaran mereka. Pemanfaatan teknologi ini memungkinkan perusahaan untuk mengumpulkan, menyimpan, dan menganalisis data historis secara lebih efektif guna memperoleh pemahaman yang lebih akurat terhadap analisis penjualan. Berdasarkan sejumlah penelitian terkini, pendekatan prediktif yang berbasis pada data historis terbukti mampu membantu perusahaan dalam memperkirakan penjualan sekaligus merancang strategi pemasaran yang lebih tepat sasaran [3].

Infinity Jar merupakan salah satu usaha berbasis online yang berfokus pada penjualan produk-produk inovatif yang dirancang untuk menarik minat konsumen. Dalam rangka memperluas jangkauan pasar serta memahami tren penjualan produknya, pelaku usaha perlu melakukan analisis terhadap produk dengan tingkat permintaan tinggi sekaligus mengidentifikasi potensi pembelian dari konsumen [4]. Melalui pemanfaatan data historis, Infinity Jar dapat mengenali produk-produk yang paling diminati, sehingga memungkinkan penyusunan strategi pemasaran yang lebih efektif dan sesuai dengan target pasar [5].

Dalam dunia pemasaran, pemanfaatan data historis memegang peranan penting dalam menyusun strategi yang lebih efektif. Melalui penerapan teknik analisis modern, pelaku usaha mampu mengenali produk-produk dengan tingkat permintaan tertinggi yang mungkin terlewatkan jika hanya menggunakan pendekatan konvensional. Sebagai contoh, analisis semacam ini dapat mengidentifikasi produk dengan potensi penjualan terbesar, sehingga memungkinkan pengalokasian sumber daya secara lebih tepat dan meningkatkan efektivitas investasi (*Return on Investment*) dalam kegiatan pemasaran [6].

Di bidang analisis data, algoritma machine learning seperti XGBoost dan Random Forest telah terbukti mampu memberikan prediksi yang akurat berdasarkan data historis. XGBoost (Extreme Gradient Boosting) sendiri mengandalkan pendekatan boosting guna meningkatkan ketepatan hasil prediksi [7]. Sementara itu, Random Forest merupakan metode *ensemble* yang mengintegrasikan sejumlah pohon keputusan untuk meningkatkan akurasi prediksi sekaligus meminimalkan risiko *overfitting* [8]. Berkat kemampuannya dalam menangani berbagai variabel yang memengaruhi hasil, kedua algoritma ini sangat cocok digunakan dalam menganalisis data transaksi penjualan yang bersifat kompleks [9].

XGBoost (Extreme Gradient Boosting) merupakan salah satu algoritma machine learning berbasis boosting yang banyak digunakan dalam analisis prediktif karena keunggulannya dalam menangani data berukuran besar, data yang tidak seimbang, serta fitur-fitur yang saling berkorelasi [10]. Dalam penerapan pada analisis penjualan, XGBoost sangat relevan karena mampu memproses data transaksi dengan variasi produk dan pola pembelian yang kompleks, serta mengungkap pola tersembunyi yang tidak teridentifikasi oleh metode konvensional [11]. Selain karena hasil akhirnya, XGBoost juga dikenal karena kecepatan dan efisiensinya dalam proses pelatihan model, sehingga menjadi pilihan tepat untuk mendukung analisis penjualan dan penyusunan strategi pemasaran yang lebih optimal di Infinity Jar.

Salah satu keunggulan utama Random Forest terletak pada tingkat interpretabilitasnya yang tinggi, karena algoritma ini membentuk sejumlah pohon keputusan secara paralel dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih stabil dan akurat [12]. Setiap

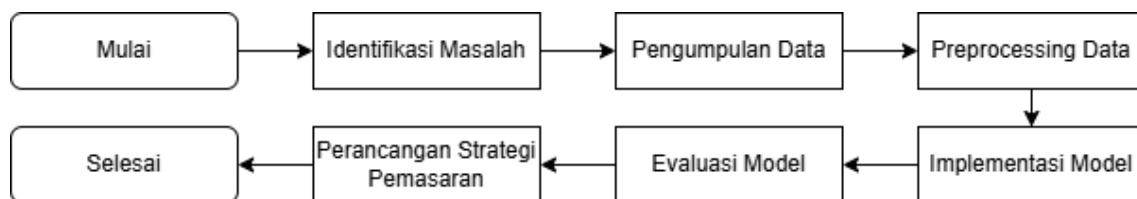
pohon dalam Random Forest mengevaluasi variabel yang berbeda-beda, sehingga memungkinkan analisis terhadap kontribusi masing-masing variabel terhadap hasil akhir [13]. Kemampuan ini sangat berguna bagi Infinity Jar, karena dapat membantu dalam memprediksi penjualan produk, seperti kategori produk maupun frekuensi pembeliannya. Dengan memahami aspek-aspek tersebut, perusahaan dapat menyesuaikan strategi pemasaran untuk meningkatkan efektivitas penjualannya.

Dengan penerapan kedua algoritma tersebut, diharapkan dapat memperoleh pemahaman yang lebih mendalam mengenai tren penjualan, termasuk memprediksi waktu pembelian konsumen, sehingga strategi pemasaran dapat disusun dengan lebih akurat dan efisien [14].

Dengan memanfaatkan XGBoost dan Random Forest untuk menganalisis data transaksi penjualan. Harapannya, penggunaan kedua algoritma ini tidak hanya dapat membantu dalam menganalisis penjualan, tetapi juga memperbaiki pengalaman konsumen. Melalui analisis data historis, pelaku usaha dapat memprediksi permintaan pembelian produk dan merancang strategi pemasaran yang lebih efektif dalam pengelolaan stok, seperti mengelompokkan produk dengan permintaan tinggi dan rendah berdasarkan prediksi penjualan.

2. METODE PENELITIAN

Dalam metode penelitian ini akan dijelaskan secara rinci setiap tahapan yang dilalui selama proses penelitian. Gambar 1 menyajikan gambaran umum alur penelitian yang menunjukkan urutan langkah-langkah yang ditempuh dalam pelaksanaan penelitian ini.



Gambar 1. Alur Penelitian

2.1. Identifikasi Masalah

Pada tahap ini, permasalahan utama yang diangkat dalam penelitian adalah bagaimana memprediksi penjualan produk secara akurat. Untuk mengidentifikasi masalah tersebut, dilakukan studi literatur melalui penelaahan berbagai artikel dan jurnal dari penelitian sebelumnya, khususnya yang membahas penerapan metode XGBoost dan Random Forest dalam kasus prediksi [15]. Studi literatur ini bertujuan untuk memahami pendekatan yang telah dilakukan dan menemukan celah atau peluang pengembangan lebih lanjut. Penelitian ini mengambil studi kasus dari sebuah toko online yang belum memiliki sistem untuk memprediksi penjualan produknya, sehingga keputusan bisnis masih bersifat manual dan kurang berbasis data.

2.2. Pengumpulan Data

Dataset yang digunakan diperoleh langsung dari toko online yang menjadi objek studi kasus. Data yang dikumpulkan mencakup informasi penjualan selama tiga tahun, yakni dari tahun 2022 hingga 2024. Dataset tersebut disusun dalam format Excel dan berisi beberapa kolom penting, antara lain tanggal transaksi, jumlah pesanan, jenis produk, kode produk, dan nomor transaksi. Informasi ini menjadi dasar untuk proses analisis dan pemodelan dalam penelitian, khususnya dalam penerapan algoritma XGBoost dan Random Forest untuk memprediksi penjualan produk.

2.3. Preprocessing Data

Proses dimulai dengan memuat dataset ke dalam lingkungan pemrograman, dilanjutkan dengan pengecekan dan penanganan missing values untuk memastikan kualitas data tetap terjaga. Setelah itu, dilakukan penyesuaian tipe data, terutama pada kolom tanggal yang dikonversi ke format datetime guna memungkinkan ekstraksi fitur waktu seperti bulan dan tahun. Tahap selanjutnya adalah pembentukan fitur baru melalui proses agregasi data, seperti pengelompokan total penjualan per produk dalam periode tertentu. Data kategorikal, seperti jenis produk, kemudian dikonversi ke dalam format numerik menggunakan metode label encoding atau one-hot encoding. Selain itu, ditambahkan fitur *lag* dengan rentang 3 periode yang merepresentasikan riwayat penjualan selama tiga periode sebelumnya. Penambahan fitur ini bertujuan agar model dapat mengenali pola historis yang relevan dalam data penjualan. Terakhir, data dibagi menjadi data latih (*train*) dan data uji (*test*) berdasarkan periode waktu yang akan diprediksi, yaitu selama satu bulan. Dengan demikian, data telah siap digunakan dalam tahap pelatihan model prediksi.

2.4. Penerapan Model

Tahap pertama dalam penerapan model adalah melakukan *hyperparameter tuning* untuk menemukan kombinasi parameter terbaik yang dapat memaksimalkan performa masing-masing model. Proses ini menggunakan metode *Grid Search*, yang dipadukan dengan teknik *cross-validation*, guna memastikan bahwa model yang dihasilkan optimal dan memiliki kestabilan yang baik. Setelah kombinasi parameter terbaik ditemukan, masing-masing model XGBoost dan Random Forest dilatih secara terpisah menggunakan data latih. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi terhadap variabel target pada data uji.

Setelah memperoleh hasil prediksi dari kedua model, langkah berikutnya adalah menggabungkan hasil prediksi tersebut menggunakan teknik *Voting Regressor*, yang merupakan metode *ensemble* dalam regresi untuk meningkatkan akurasi dan stabilitas hasil prediksi secara keseluruhan. Pada teknik ini, nilai prediksi akhir dihitung dengan cara melakukan *averaging* atau rata-rata dari hasil prediksi yang dihasilkan oleh masing-masing model. Teknik ini dikenal sebagai *soft voting*, di mana prediksi akhir diperoleh dengan rumus:

$$y_{Voting} = \frac{y_{XGBoost} + y_{Random Forest}}{2} \quad (1)$$

Keterangan:

y_{Voting} = Hasil akhir *Voting Regressor*
 $y_{XGBoost}$ = Hasil prediksi dari XGBoost
 $y_{Random Forest}$ = Hasil prediksi dari Random Forest

Rumus pada persamaan 1 tersebut digunakan dalam teknik *Voting Regressor* yang menggabungkan hasil prediksi dari dua model regresi, di mana nilai prediksi akhir diperoleh dengan cara menghitung rata-rata dari hasil prediksi kedua model. Dengan menggunakan dua model, pembagi dalam rumus adalah angka 2, menandakan jumlah model yang digabungkan.

Pemilihan algoritma XGBoost dan Random Forest didasarkan pada keunggulan keduanya dalam membangun model prediksi regresi yang akurat dan andal. XGBoost dikenal sebagai algoritma *boosting* yang efisien dan mampu menangani kompleksitas data melalui regularisasi, sedangkan Random Forest unggul dalam stabilitas prediksi. Kombinasi keduanya dalam metode *ensemble* diharapkan mampu meningkatkan performa model dengan menggabungkan kelebihan masing-masing model dasar.

2.5. Evaluasi Model

Setelah proses pelatihan dan prediksi dilakukan menggunakan model hybrid, tahap selanjutnya adalah mengevaluasi performa model. Evaluasi dilakukan dengan membandingkan hasil prediksi dengan data aktual menggunakan tiga metrik evaluasi regresi, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Mean Absolute Percentage Error* (MAPE). MAE mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, sehingga semakin kecil nilainya, semakin akurat model dalam memprediksi. MSE menghitung rata-rata kuadrat dari selisih antara nilai prediksi dan aktual, dan memberikan penalti lebih besar terhadap kesalahan yang besar. Sementara itu, MAPE menyatakan rata-rata kesalahan dalam bentuk persentase terhadap nilai aktual, sehingga memudahkan interpretasi seberapa besar kesalahan relatif yang terjadi. Nilai MAE, MSE, dan MAPE yang rendah menunjukkan bahwa model memiliki performa yang baik dan mampu memetakan pola penjualan dengan efektif dan andal.

2.6. Perancangan Strategi Pemasaran

Berdasarkan hasil prediksi total penjualan, produk diklasifikasikan menjadi dua kategori, yaitu permintaan tinggi dan permintaan rendah. Klasifikasi ini dilakukan dengan menggunakan nilai median dari hasil prediksi sebagai batas pemisah. Produk dengan nilai prediksi penjualan di atas median dikategorikan sebagai produk berpermintaan tinggi, sementara produk di bawah median termasuk dalam kategori permintaan rendah. Strategi pemasaran kemudian dirancang berdasarkan klasifikasi tersebut. Untuk produk dengan permintaan tinggi, disarankan dilakukan penambahan stok guna mengantisipasi tingginya permintaan pasar. Sebaliknya, produk dengan permintaan rendah disarankan untuk dikurangi stoknya agar tidak terjadi penumpukan yang berpotensi menimbulkan kerugian. Selain itu, diterapkan pula strategi bundling, yaitu menggabungkan satu produk dengan permintaan tinggi dan satu produk dengan permintaan rendah ke dalam satu paket penjualan. Strategi ini bertujuan untuk meningkatkan daya tarik terhadap produk dengan permintaan rendah sekaligus mendorong peningkatan total penjualan.

3. HASIL DAN PEMBAHASAN

3.1. Penerapan Model

Tahap penerapan model dimulai dengan proses pelatihan terhadap masing-masing model individu, yaitu XGBoost dan Random Forest, yang dilakukan secara terpisah menggunakan data latih. Rincian dari masing-masing parameter model yang digunakan dijelaskan sebagai berikut:

1. XGBoost

Penerapan model XGBoost dimulai dengan proses *hyperparameter tuning* untuk menentukan kombinasi parameter yang paling optimal dalam meningkatkan kinerja model. Skema *tuning* tersebut dapat dilihat pada Tabel 1, yang mencantumkan berbagai rentang nilai untuk setiap parameter.

Tabel 1. Skema *Hyperparameter Tuning* XGBoost

Parameter	Rentang Nilai
Jumlah Estimators	5, 50, 100
Max Depth	5, 10, 15
Learning Rate	0,05, 0,1, 0,3
Subsample	0,8, 1,0
Colsample Bytree	0,8, 1,0

Setelah melalui proses pencarian parameter terbaik, diperoleh konfigurasi optimal yang digunakan dalam pelatihan akhir model, sebagaimana ditampilkan pada Tabel 2.

Tabel 2. Hasil Akhir Parameter XGBoost

Parameter	Rentang Nilai
Jumlah Estimators	100
Max Depth	15
Learning Rate	0,1
Subsample	0,8
Colsample Bytree	0,8

2. Random Forest

Penerapan model Random Forest dimulai dengan proses *hyperparameter tuning* guna menemukan kombinasi parameter terbaik yang mampu meningkatkan performa model secara maksimal. Skema tuning yang digunakan dapat dilihat pada Tabel 3, yang memuat rentang nilai untuk setiap parameter yang diuji.

Tabel 3. Skema Hyperparameter Tuning Random Forest

Parameter	Rentang Nilai
Jumlah Estimators	50, 100, 200
Max Depth	5, 10, 15, None
Min Samples Split	2, 5
Min Samples Leaf	1, 2

Setelah proses tuning dilakukan, diperoleh kombinasi parameter terbaik yang digunakan dalam pelatihan akhir model Random Forest. Detail konfigurasi parameter optimal tersebut disajikan pada Tabel 4.

Tabel 4. Hasil Akhir Parameter Random Forest

Parameter	Rentang Nilai
Jumlah Estimators	50
Max Depth	10
Min Samples Split	5
Min Samples Leaf	2

Setelah diperoleh kombinasi parameter terbaik untuk masing-masing model, selanjutnya melatih model XGBoost dan Random Forest secara terpisah menggunakan konfigurasi tersebut. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi pada data uji. Kemudian membentuk model *hybrid* dengan menggunakan *Voting Regressor* guna menghasilkan prediksi akhir yang lebih akurat dan stabil. Pendekatan yang digunakan adalah *soft voting*, yaitu dengan menghitung rata-rata (*average*) dari hasil prediksi model XGBoost dan Random Forest.

3.2. Evaluasi Model

Tabel 5 menyajikan hasil performa dari tiga model prediktif yang digunakan, yaitu XGBoost, Random Forest, dan model *hybrid* XGBoost–Random Forest. Evaluasi dilakukan

menggunakan tiga metrik utama, yakni *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE).

Tabel 5. Hasil Evaluasi Model

Model	MSE	MAE	MAPE
XGBoost	6,91	1,27	33,01%
Random Forest	5,45	1,09	23,48%
XGBoost - Random Forest	5,36	1,00	14,39%

Berdasarkan hasil evaluasi, model hybrid XGBoost–Random Forest menunjukkan performa terbaik dibandingkan kedua model individu sebelum dilakukan proses penggabungan melalui *Voting Regressor*. Model *hybrid* ini menghasilkan nilai *Mean Squared Error* (MSE) sebesar 5,36, *Mean Absolute Error* (MAE) sebesar 1,00, dan *Mean Absolute Percentage Error* (MAPE) sebesar 14,39%.

Nilai MAPE sebesar 14,39% mengindikasikan bahwa model *hybrid* memiliki tingkat akurasi yang baik dan cukup andal dalam melakukan prediksi. Hasil ini membuktikan bahwa kombinasi kedua model melalui pendekatan *ensemble learning* secara signifikan mampu mengurangi tingkat kesalahan prediksi dibandingkan penggunaan model tunggal. Dengan demikian, penerapan *Voting Regressor* terbukti efektif dalam meningkatkan kinerja model prediksi berbasis data penjualan.

3.3. Perancangan Strategi Pemasaran

Hasil prediksi menunjukkan bahwa terdapat 11 produk yang diprediksi penjualannya. Produk-produk tersebut kemudian diklasifikasikan berdasarkan nilai median penjualan ke dalam dua kategori permintaan, yaitu permintaan tinggi dan permintaan rendah. Dari hasil klasifikasi, diperoleh 2 produk yang termasuk dalam kategori permintaan tinggi dan 9 produk dalam kategori permintaan rendah.

Sebagai langkah lanjutan, diterapkan strategi pemasaran berupa *product bundling*, yaitu dengan menggabungkan produk dari kategori permintaan tinggi dengan produk dari kategori permintaan rendah. Hasil dari proses *bundling* ini menghasilkan 18 paket produk yang dapat dijadikan pertimbangan dalam perencanaan.

Selain itu, sebagai bagian dari rekomendasi strategi pemasaran juga, disarankan agar permintaan terhadap produk dalam kategori rendah dikendalikan atau dikurangi untuk menghindari potensi penumpukan stok (*overstocking*) yang dapat berdampak negatif pada efisiensi operasional dan keuangan. Sebaliknya, produk dalam kategori permintaan tinggi dianjurkan untuk ditingkatkan stok dan ketersediaannya agar dapat memenuhi potensi permintaan yang besar dan mencegah kekosongan barang (*stockout*).

4. KESIMPULAN

Berdasarkan hasil penelitian, penerapan algoritma XGBoost dan Random Forest yang digabungkan melalui metode *Voting Regressor* terbukti mampu meningkatkan akurasi prediksi penjualan produk, dengan nilai evaluasi terbaik yaitu MSE sebesar 5,36, MAE sebesar 1,00, dan MAPE sebesar 14,39%. Model *hybrid* ini lebih unggul dibandingkan model tunggal, di mana Random Forest memiliki performa lebih baik (MAPE 23,48%) dibandingkan XGBoost (MAPE 33,01%). Selain itu, hasil prediksi yang diperoleh digunakan untuk menyusun strategi *bundling* produk dan pengelolaan stok, yang menghasilkan 18 kombinasi paket produk berdasarkan kategori permintaan tinggi dan rendah.

5. SARAN

Beberapa usulan yang dapat dipertimbangkan untuk pengembangan penelitian ini di masa mendatang adalah sebagai berikut: 1) Mengeksplorasi metode *ensemble learning* lainnya untuk menentukan pendekatan penggabungan model yang paling optimal. 2) Menambahkan dataset dengan rentang waktu yang lebih panjang untuk meningkatkan akurasi prediksi dan memperoleh tren penjualan yang lebih representatif secara jangka panjang. 3) Mengintegrasikan algoritma berbasis time series seperti ARIMA, Prophet, atau LSTM untuk memahami pola musiman dan tren penjualan, serta mengevaluasi pengaruh variabel waktu terhadap hasil prediksi.

DAFTAR PUSTAKA

- [1] Yolanda, F., & Dwiridhotjahjono, J. (2021). Strategi Pemasaran dalam Meningkatkan Penjualan Antena HDF Surabaya. *Value : Jurnal Manajemen Dan Akuntansi*, 16(1), 231–239. <https://doi.org/10.32534/jv.v16i1.1841>
- [2] Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122–137. <https://doi.org/10.34010/jati.v13i2.10090>
- [3] Alghazzawi, D. M., Alquraishee, A. G. A., Badri, S. K., & Hasan, S. H. (2023). ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review. *Sustainability (Switzerland)*, 15(9). <https://doi.org/10.3390/su15097076>
- [4] Wahyu Hadi Sutiyono, & Widya Setiafindari. (2024). Analisis Penerapan Forecasting Penjualan Untuk Menentukan Jumlah Tenaga Kerja Efektif Produksi Tepung Mocaf Pada UMKM XYZ. *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro Dan Informatika*, 2(4), 181–194. <https://doi.org/10.61132/jupiter.v2i4.423>
- [5] Massaro, A., Panarese, A., Giannone, D., & Galiano, A. (2021). Augmented data and Xgboost improvement for sales forecasting in the large-scale retail sector. *Applied Sciences (Switzerland)*, 11(17). <https://doi.org/10.3390/app11177793>
- [6] Wijaya, H., Hostiadi, D. P., & Triandini, E. (2024). Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi Random Search Pada Model Xgboost. *SPINTER (Prosiding Seminar Hasil Penelitian Informatika Dan Komputer)*, 1(2), 829–833.
- [7] Jabeur, S. Ben, Mefteh-Wali, S., & Viviani, J. L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 334(1–3), 679–699. <https://doi.org/10.1007/s10479-021-04187-w>
- [8] Putra, M. R. P. (2024). Model Pembelajaran Mesin Hibridisasi RFXGB Untuk Prediksi Drop Out Siswa MOOC. *G-Tech: Jurnal Teknologi Terapan*, 8, 2485–2494. <https://doi.org/10.70609/gtech.v8i4.5179>
- [9] Chen, H. (2023). Enterprise marketing strategy using big data mining technology combined with XGBoost model in the new economic era. *PLoS ONE*, 18(6 June), 1–22. <https://doi.org/10.1371/journal.pone.0285506>
- [10] Siringoringo, R., Perangin-angin, R., & Purba, M. J. (2021). Segmentasi Dan Peramalan Pasar Retail Menggunakan Xgboost Dan Principal Component Analysis. *METHOMIKA*

Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi, 5(1), 42–47.
<https://doi.org/10.46880/jmika.vol5no1.pp42-47>

- [11] Li, K. (2023). A Sales Prediction Method Based on XGBoost Algorithm Model. *BCP Business & Management*, 36, 367–371. <https://doi.org/10.54691/bcpbm.v36i.3487>
- [12] Yang, T. (2023). Sales Prediction of Walmart Sales Based on OLS, Random Forest, and XGBoost Models. *Highlights in Science, Engineering and Technology*, 49, 244–249. <https://doi.org/10.54097/hset.v49i.8513>
- [13] Zhou, K., Li, X., Sun, T., & Luo, B. (2023). Research on Vegetable Sales Prediction and Strategy of Supermarkets Based on XGBoost and Random Forest Modeling. *Proceedings - 2023 International Conference on Applied Physics and Computing, ICAPC 2023*, 373–378. <https://doi.org/10.1109/ICAPC61546.2023.00076>
- [14] Ashraf, D. M. U. (2022). A Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, 6(04), 23–33. <https://doi.org/10.54692/lgurjcsit.2022.0604399>
- [15] Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 111–119. <https://doi.org/10.37148/bios.v5i2.128>