

Prediksi Permintaan Layanan Percetakan Menggunakan Algoritma Random Forest Berdasarkan Data Historis Transaksi

Farid Anwar Shidiq^{1*}, Hanifah Permatasari², Nibras Faiq Muhammad³

^{1*,2,3}Teknik Informatika, Universitas Duta Bangsa Surakarta

*Email: faridanwar252@gmail.com

Abstract

Fluctuating demand for printing services causes difficulties in operational planning, such as raw material procurement, workforce scheduling, and service capacity management. Therefore, this study aims to build a prediction model for printing service demand using the Random Forest Regression algorithm based on historical transaction data. The study was conducted using the CRISP-DM approach which includes the stages of Business understanding, Data Understanding, Data Preparation, Modeling, and Evaluation. In the data preparation stage, feature engineering was carried out by forming calendar attributes and transaction history attributes to improve the model's ability to recognize demand patterns. Next, hyperparameter optimization was carried out before the model training process. Model evaluation was carried out using the TimeSeriesSplit method with five folds using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-Squared (R^2) metrics. The results showed that the Random Forest Regression model obtained an average MAE value of 39.20, RMSE of 63.62, and R^2 of 0,7017. These results indicate that the model is capable of predicting demand for printing services based on historical transaction data, although there are still opportunities to improve model performance by adding more relevant variables and using longer historical data.

Keywords: Demand prediction, random forest regression, crisp-dm, timeseriessplit, feature engineering

Abstrak

Permintaan layanan percetakan yang berfluktuasi menyebabkan kesulitan dalam perencanaan operasional, seperti penyediaan bahan baku, penjadwalan tenaga kerja, dan pengelolaan kapasitas layanan. Oleh karena itu, penelitian ini bertujuan untuk membangun model prediksi permintaan layanan percetakan menggunakan algoritma *Random Forest Regression* berdasarkan data historis transaksi. Penelitian dilakukan dengan pendekatan CRISP-DM yang meliputi tahapan *Business understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*. Pada tahap data preparation dilakukan *feature engineering* dengan membentuk atribut kalender dan atribut histori transaksi untuk meningkatkan kemampuan model dalam mengenali pola permintaan. Selanjutnya dilakukan optimasi hyperparameter sebelum proses pelatihan model. Evaluasi model dilakukan menggunakan metode *TimeSeriesSplit* sebanyak lima fold dengan metrik Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-Squared (R^2). Hasil penelitian menunjukkan bahwa model *Random Forest Regression* memperoleh nilai rata-rata MAE sebesar 39,20, RMSE sebesar 63,62, dan R^2 sebesar 0,7017. Hasil tersebut menunjukkan bahwa model mampu melakukan prediksi permintaan layanan percetakan berdasarkan data historis transaksi, meskipun masih terdapat peluang untuk meningkatkan performa model melalui penambahan variabel yang lebih relevan dan penggunaan data historis yang lebih panjang.

Kata kunci: Prediksi permintaan, random forest regression, crisp-dm, timeseriessplit, feature engineering.

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong pemanfaatan data dalam berbagai sektor usaha untuk mendukung proses pengambilan keputusan secara lebih efektif dan efisien. Salah satu teknologi yang banyak diterapkan dalam pengolahan data adalah *machine learning*, yaitu metode yang memungkinkan sistem melakukan analisis data historis untuk menemukan pola tertentu dan menghasilkan prediksi berdasarkan data sebelumnya. Dalam dunia bisnis, penerapan *machine learning* banyak digunakan untuk melakukan analisis perilaku pelanggan, prediksi penjualan, hingga prediksi permintaan produk maupun layanan guna meningkatkan efisiensi operasional perusahaan [1]. Data transaksi yang tersimpan dalam aktivitas bisnis memiliki potensi besar untuk dimanfaatkan sebagai sumber informasi dalam menentukan strategi operasional dan pengelolaan sumber daya secara lebih optimal.

Pada sektor usaha jasa, khususnya usaha percetakan, permintaan layanan memiliki karakteristik yang dinamis dan fluktuatif. Permintaan layanan percetakan dapat mengalami peningkatan pada periode tertentu, seperti kegiatan akademik, administrasi, maupun kebutuhan organisasi masyarakat. Kondisi tersebut sering menyebabkan pelaku usaha mengalami kesulitan dalam memperkirakan jumlah permintaan layanan pada waktu tertentu. Ketidakmampuan dalam memprediksi permintaan secara akurat dapat berdampak pada kurang optimalnya pengelolaan bahan baku, waktu operasional, serta kualitas pelayanan kepada pelanggan. Oleh karena itu, diperlukan suatu metode analisis yang mampu memanfaatkan data historis transaksi untuk menghasilkan prediksi permintaan layanan secara lebih akurat [2].

ADIA SOLUTION merupakan salah satu usaha percetakan yang memiliki data historis transaksi layanan percetakan, seperti print, fotokopi, dan cetak foto. Data transaksi tersebut menyimpan informasi penting mengenai pola permintaan pelanggan yang dapat dimanfaatkan untuk proses analisis dan prediksi permintaan layanan pada periode mendatang. Pemanfaatan data historis transaksi dalam proses prediksi dapat membantu pelaku usaha dalam melakukan perencanaan operasional, persiapan bahan baku, pengaturan waktu kerja, serta meningkatkan efektivitas pelayanan kepada pelanggan [3]. Dengan adanya sistem prediksi permintaan layanan, pelaku usaha dapat mengambil keputusan secara lebih tepat berdasarkan pola data yang telah dianalisis.

Salah satu algoritma *machine learning* yang memiliki performa baik dalam proses prediksi adalah Random Forest. Algoritma ini termasuk dalam metode *ensemble learning* yang bekerja dengan membangun sejumlah *decision tree* dan menggabungkan hasil prediksi dari setiap pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat. Random Forest memiliki beberapa keunggulan, seperti mampu menangani data dalam jumlah besar, mengurangi risiko *overfitting*, serta menghasilkan tingkat performa model yang baik pada berbagai kasus prediksi [4]. Penelitian sebelumnya menunjukkan bahwa algoritma *Random Forest Regression* mampu menghasilkan performa prediksi yang baik dalam kasus prediksi penjualan berbasis data historis transaksi [5]. Selain itu, metode ini juga dinilai efektif dalam menangani pola data yang kompleks dan bervariasi.

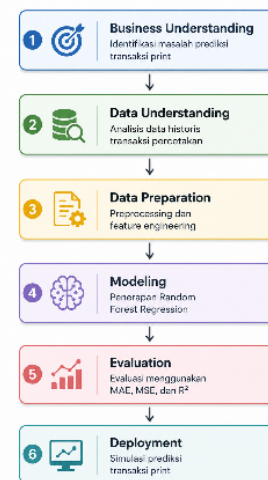
Beberapa penelitian terdahulu telah menerapkan algoritma Random Forest dalam bidang prediksi penjualan dan analisis data

transaksi. Penelitian yang dilakukan oleh Hidayat et al. menunjukkan bahwa *Random Forest Regression* mampu menghasilkan model prediksi penjualan dengan tingkat kinerja model yang baik pada sektor ritel [6]. Penelitian lain yang dilakukan oleh Tantyoko et al. juga menunjukkan bahwa metode *Random Forest* memiliki performa yang baik dalam proses prediksi karena mampu mengurangi *overfitting* dan meningkatkan kinerja model [5]. Namun demikian, penelitian yang secara khusus membahas penerapan algoritma *Random Forest Regression* untuk memprediksi permintaan layanan percetakan berdasarkan data historis transaksi masih relatif terbatas. Sebagian besar penelitian sebelumnya lebih berfokus pada prediksi penjualan produk pada sektor ritel maupun analisis pola transaksi menggunakan metode asosiasi.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model prediksi permintaan layanan percetakan menggunakan algoritma *Random Forest Regression* berdasarkan data historis transaksi pada ADIA SOLUTION. Proses pengembangan model dilakukan menggunakan pendekatan *Cross Industry Standard Process for Data Mining (CRISP-DM)* yang terdiri dari tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Evaluasi performa model dilakukan menggunakan metode *TimeSeriesSplit* sebanyak lima *fold* dengan metrik *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-Squared (R²)* untuk mengukur performa model prediksi yang dihasilkan. Penelitian ini diharapkan dapat memberikan kontribusi dalam penerapan *machine learning* pada bidang usaha percetakan serta membantu pelaku usaha dalam melakukan perencanaan operasional dan pengambilan keputusan secara lebih efektif.

2. Metoda Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan metode eksperimen untuk membangun model prediksi permintaan layanan percetakan berbasis data historis transaksi pada ADIA SOLUTION. Dataset yang digunakan berasal dari transaksi layanan percetakan, seperti print, fotokopi, dan cetak foto, yang dikumpulkan dalam periode tertentu. Penelitian dilakukan menggunakan pendekatan *Cross Industry Standard Process for Data Mining (CRISP-DM)* karena mampu memberikan alur pengolahan data yang sistematis mulai dari identifikasi masalah hingga implementasi model prediksi.



Gambar 1. Diagram Alir Proses CRISP-DM

Tahap awal penelitian dimulai dengan *business understanding* untuk mengidentifikasi permasalahan utama pada usaha percetakan, yaitu ketidakstabilan permintaan layanan yang menyulitkan proses perencanaan operasional. Selanjutnya, tahap *data understanding* dilakukan melalui pengumpulan dan analisis awal dataset guna mengetahui karakteristik, struktur, serta pola data transaksi yang tersedia. Setelah itu, tahap *data preparation* dilakukan dengan membersihkan data, menangani *missing value*, serta melakukan transformasi fitur berdasarkan waktu transaksi agar data siap digunakan dalam proses pelatihan model.

Dataset yang telah melalui proses data preparation selanjutnya digunakan dalam proses pembentukan model menggunakan algoritma *Random Forest Regression*. Evaluasi model dilakukan menggunakan metode *TimeSeriesSplit* sebanyak lima *fold*, sehingga pada setiap iterasi data historis digunakan sebagai data pelatihan (*training data*), sedangkan data pada periode berikutnya digunakan sebagai data pengujian (*testing data*). Pendekatan ini dipilih untuk mempertahankan urutan kronologis data transaksi sehingga lebih sesuai untuk permasalahan prediksi berbasis deret waktu (*time series*).

Proses modeling dilakukan menggunakan algoritma *Random Forest Regression* yang diimplementasikan melalui bahasa pemrograman *Python* dengan bantuan *library scikit-learn*. Algoritma ini dipilih karena mampu menangani data yang kompleks dan menghasilkan prediksi yang stabil melalui kombinasi beberapa *decision tree*. Kinerja model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* untuk mengukur tingkat kemampuan prediksi model. Tahap akhir berupa *deployment* dilakukan dengan menerapkan model yang telah dibangun sebagai alat bantu dalam memperkirakan permintaan layanan percetakan sehingga dapat mendukung pengambilan keputusan operasional secara lebih efektif.

3. Hasil Penelitian

3.1. Business Understanding

Tahap *Business understanding* dilakukan untuk mengidentifikasi permasalahan yang dihadapi oleh ADIA SOLUTION dalam mengelola layanan percetakan. Berdasarkan hasil observasi dan analisis proses bisnis, permintaan layanan print memiliki pola yang fluktuatif sehingga sulit diperkirakan secara manual. Kondisi tersebut menyebabkan pelaku usaha mengalami kesulitan dalam melakukan perencanaan operasional, seperti persiapan bahan baku, pengaturan waktu kerja, serta pengelolaan layanan kepada pelanggan.

Untuk mengatasi permasalahan tersebut, diperlukan suatu model prediksi yang mampu memperkirakan jumlah permintaan layanan print berdasarkan data historis transaksi. Prediksi yang dihasilkan diharapkan dapat membantu pelaku usaha dalam mengambil keputusan operasional secara lebih efektif dan efisien. Oleh karena itu, penelitian ini menerapkan algoritma *Random Forest Regression* untuk membangun model prediksi permintaan layanan percetakan menggunakan data historis transaksi pada ADIA SOLUTION.

3.2. Data Understanding

Dataset yang digunakan dalam penelitian ini merupakan data historis transaksi harian layanan percetakan pada ADIA SOLUTION yang diperoleh dari aktivitas operasional perusahaan. Data transaksi mencakup berbagai jenis layanan seperti layanan percetakan, layanan ATK, dan layanan lainnya. Seluruh data transaksi disimpan dalam format Microsoft Excel dan kemudian diolah menggunakan bahasa pemrograman *Python* sebagai dasar pembangunan model prediksi. Pendekatan analisis data transaksi pelanggan semacam ini sejalan dengan penelitian yang menerapkan proses Knowledge Discovery in Database untuk mengidentifikasi pola perilaku pelanggan pada sektor telekomunikasi guna mendukung pengambilan keputusan berbasis data [7].

Pada tahap awal, seluruh data transaksi digabungkan menjadi dataset harian berdasarkan tanggal transaksi. Selanjutnya dilakukan proses *preprocessing* untuk menghasilkan pemodelan dataset yang telah dilengkapi dengan atribut kalender dan fitur histori transaksi. Setelah proses *preprocessing* selesai, diperoleh sebanyak 515 data transaksi harian yang siap digunakan dalam proses pelatihan dan evaluasi model.

Tabel 1. Atribut Dataset Penelitian

Atribut	Keterangan
Waktu	Tanggal, Hari, Bulan, Minggu, HariDalamMinggu, Weekend
Layanan	total_ATK, total_Non,

Atribut	Keterangan
Target	total_Percetakan, total_semua Percetakan

Dataset penelitian terdiri atas atribut waktu, atribut layanan, feature histori, dan satu variabel target. Atribut waktu digunakan untuk merepresentasikan karakteristik temporal transaksi, sedangkan atribut layanan menggambarkan jumlah transaksi pada setiap kategori layanan percetakan. Selain itu, feature histori dibentuk melalui proses *feature engineering* untuk membantu model mengenali pola permintaan berdasarkan transaksi pada periode sebelumnya. Variabel target pada penelitian ini adalah jumlah transaksi layanan print harian.

3.3. Data Preparation

Tahap data *preparation* dilakukan untuk mempersiapkan dataset sebelum digunakan pada proses pelatihan model *Random Forest Regression*. Tahap ini meliputi proses *preprocessing* dan *feature engineering* menggunakan bahasa pemrograman *Python* dengan pustaka *pandas*. Proses diawali dengan mengonversi atribut tanggal ke dalam format *datetime*, kemudian mengurutkan data berdasarkan urutan waktu transaksi agar pola historis tetap terjaga.

Selanjutnya dilakukan pembentukan atribut kalender yang terdiri atas Hari, Bulan, Minggu, HariDalamMinggu, dan Weekend. Atribut tersebut digunakan untuk merepresentasikan karakteristik waktu yang diduga memengaruhi pola permintaan layanan percetakan. Pendekatan pemanfaatan atribut kalender untuk memprediksi pola permintaan berbasis waktu ini juga terbukti efektif diterapkan pada algoritma Random Forest dalam memprediksi volume kunjungan pada sektor layanan publik hanya berdasarkan sinyal kalender [8].

Selain atribut kalender, penelitian ini menerapkan *feature engineering* berbasis histori transaksi print untuk membantu model mempelajari pola permintaan pada periode sebelumnya. Fitur yang dibentuk meliputi lag

feature, moving average, rolling statistic, serta fitur interaksi yang dihasilkan dari kombinasi beberapa atribut histori transaksi. Setelah seluruh fitur berhasil dibentuk, data yang masih memiliki nilai kosong akibat proses shift dan rolling window dihapus sehingga diperoleh dataset yang siap digunakan pada tahap pemodelan. Penggunaan *lag feature* dan *rolling statistic* sebagai representasi pola historis pada data deret waktu juga diterapkan pada penelitian yang menganalisis pengaruh rekayasa fitur berbasis *rolling statistics* terhadap performa model Random Forest dan *Gradient Boosting* dalam prediksi data *time series* [9].

Tabel 2. *Feature Engineering*

Fitur	Deskripsi
Feature Kalender	day, month, year, week, dayofweek, weekend, quarter, dayofyear, serta fitur siklus waktu (day_sin, day_cos, month_sin, month_cos, dow_sin, dow_cos, week_sin, week_cos, doy_sin, doy_cos)
Lag Feature	lag_1, lag_2, lag_3, lag_5, lag_7, lag_10, lag_14, lag_21, lag_30, lag_45, lag_60
Rolling Feature	rolling_mean, rolling_std, rolling_max, rolling_min, rolling_trend
Feature Permintaan	high_demand
Feature Interaksi	max_x_high, volatility

Penambahan fitur-fitur tersebut bertujuan agar model *Random Forest Regression* tidak hanya memanfaatkan informasi transaksi pada hari pengamatan, tetapi juga mampu mengenali pola historis, kecenderungan permintaan, serta variasi transaksi yang terjadi pada periode sebelumnya. Proses *feature engineering* menghasilkan 117 atribut baru, sehingga jumlah atribut pada dataset meningkat dari 6 atribut menjadi 123 atribut. Setelah seluruh fitur berhasil dibentuk, data yang masih memiliki nilai kosong akibat proses lag dan rolling window dihapus sehingga diperoleh sebanyak 512 data yang siap digunakan pada tahap optimasi hyperparameter dan pemodelan.

3.4. Optimasi Hyperparameter

Sebelum membangun model *Random Forest Regression*, dilakukan proses optimasi *hyperparameter* untuk memperoleh kombinasi parameter yang menghasilkan performa prediksi terbaik. Optimasi dilakukan secara bertahap menggunakan bahasa pemrograman Python dengan pustaka *scikit-learn*. Setiap parameter diuji menggunakan metode *TimeSeriesSplit* sebanyak lima fold sehingga proses evaluasi tetap mempertahankan urutan waktu pada data historis transaksi. Pendekatan optimasi *hyperparameter* menggunakan metode *grid search* untuk meningkatkan akurasi model prediksi berbasis data deret waktu juga diterapkan pada penelitian peramalan tingkat inflasi menggunakan pendekatan *Ensemble Bagging* pada algoritma *Support Vector Regression* [10].

Parameter yang dioptimasi meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_features*, dan *bootstrap*. Setiap kombinasi parameter dievaluasi menggunakan nilai *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2). Kombinasi parameter yang menghasilkan nilai R^2 tertinggi dipilih sebagai parameter terbaik untuk digunakan pada proses pembentukan model akhir.

Tabel 3. Parameter yang Dioptimasi

Parameter	Nilai
<i>n_estimators</i>	100, 200, 300, 500, 700, 1000
<i>max_depth</i>	10, 15, 20, 25, 30, None
<i>min_samples_split</i>	2, 5, 10, 15
<i>min_samples_leaf</i>	1, 2, 4
<i>max_features</i>	<i>sqrt</i> , <i>log2</i> , None
<i>bootstrap</i>	True, False

Berdasarkan proses optimasi, diperoleh kombinasi parameter terbaik yang selanjutnya digunakan pada proses pelatihan model *Random Forest Regression*.

Tabel 4. Hasil Optimasi Hyperparameter

Parameter	Nilai
<i>n_estimators</i>	500
<i>max_depth</i>	15

<i>min_samples_split</i>	2
<i>min_samples_leaf</i>	1
<i>max_features</i>	<i>sqrt</i>
<i>bootstrap</i>	False

3.5. Modeling

Setelah diperoleh kombinasi *hyperparameter* terbaik pada tahap optimasi, proses selanjutnya adalah membangun model prediksi menggunakan algoritma *Random Forest Regression*. Model dibangun menggunakan parameter hasil optimasi sehingga konfigurasi yang digunakan telah disesuaikan dengan karakteristik dataset penelitian. Tahap ini bertujuan untuk menghasilkan model yang mampu memprediksi jumlah permintaan layanan print berdasarkan data historis transaksi.

Variabel masukan (*input*) pada penelitian ini terdiri atas atribut kalender, atribut layanan, serta fitur histori transaksi yang dihasilkan pada tahap *feature engineering*. Variabel target (*output*) yang diprediksi adalah jumlah transaksi layanan percetakan harian.

Proses pelatihan dan pengujian model dilakukan menggunakan metode *TimeSeriesSplit* sebanyak lima fold. Pada setiap fold, model dilatih menggunakan data historis yang tersedia, kemudian diuji pada data transaksi berikutnya sesuai urutan waktu. Pendekatan ini dipilih karena mampu mempertahankan urutan kronologis data sehingga lebih sesuai digunakan pada permasalahan prediksi berbasis deret waktu (*time series*).

Tabel 5. Parameter *Random Forest Regression*

Parameter	Nilai
<i>n_estimators</i>	500
<i>max_depth</i>	15
<i>min_samples_split</i>	2
<i>min_samples_leaf</i>	1
<i>max_features</i>	<i>sqrt</i>
<i>bootstrap</i>	True

Selama proses pelatihan, setiap *decision tree* dibangun menggunakan sebagian data

hasil *bootstrap sampling* dan sebagian fitur yang dipilih secara acak. Prediksi akhir diperoleh melalui proses agregasi berupa rata-rata hasil prediksi seluruh *decision tree*, sehingga menghasilkan nilai prediksi yang lebih stabil dibandingkan penggunaan satu pohon keputusan tunggal. Pendekatan berbasis *Knowledge Discovery in Database* dengan algoritma *ensemble* seperti *Gradient Boosting* juga telah terbukti mampu menghasilkan model prediksi dengan akurasi tinggi dalam mengklasifikasikan tingkat keparahan berdasarkan pola data yang kompleks [11].

3.6. Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur kemampuan model *Random Forest Regression* dalam memprediksi jumlah permintaan layanan print berdasarkan data historis transaksi. Evaluasi model dilakukan menggunakan metode *TimeSeriesSplit* sebanyak lima *fold* sehingga proses pelatihan dan pengujian dilakukan secara bertahap mengikuti urutan waktu transaksi. Pendekatan ini dipilih untuk menjaga karakteristik data deret waktu (*time series*) sehingga model selalu dilatih menggunakan data historis dan diuji pada data yang terjadi setelahnya.

Pada setiap *fold*, model *Random Forest Regression* dibangun menggunakan data pelatihan, kemudian menghasilkan prediksi terhadap data pengujian. Nilai prediksi dibandingkan dengan data aktual untuk menghitung nilai *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2). Setelah seluruh *fold* selesai dievaluasi, nilai MAE, RMSE, dan R^2 dihitung rata-ratanya sehingga diperoleh gambaran performa model secara keseluruhan.

Tabel 6. Hasil Evaluasi Model

Metrik Evaluasi	Nilai
MAE Rata-rata	39,20
RMSE Rata-rata	63,62
R^2 Rata-rata	0,7071

Berdasarkan hasil evaluasi pada Tabel 6, model *Random Forest Regression* memperoleh

nilai MAE rata-rata sebesar 39,20. Nilai tersebut menunjukkan bahwa rata-rata selisih absolut antara hasil prediksi dan nilai aktual adalah sekitar 39 transaksi. Semakin kecil nilai MAE menunjukkan bahwa hasil prediksi model semakin mendekati data aktual.

Selain itu, model memperoleh nilai RMSE rata-rata sebesar 63,62. Nilai RMSE memberikan gambaran besarnya kesalahan prediksi dengan memberikan penalti yang lebih besar terhadap kesalahan yang tinggi. Oleh karena itu, RMSE digunakan sebagai pelengkap MAE untuk mengevaluasi kestabilan performa model selama proses *cross validation*.

Nilai *R-Squared* (R^2) rata-rata sebesar 0,7071 menunjukkan bahwa model mampu menjelaskan sekitar 70,71% variasi permintaan layanan print berdasarkan variabel-variabel yang digunakan dalam penelitian. Hasil tersebut menunjukkan bahwa model telah mampu menangkap sebagian pola historis transaksi, namun masih terdapat variasi permintaan yang dipengaruhi oleh faktor lain di luar variabel penelitian, seperti kondisi akademik, hari libur, promosi, maupun faktor operasional lainnya.

Nilai R^2 yang masih relatif rendah menunjukkan bahwa permintaan layanan percetakan juga dipengaruhi oleh faktor-faktor lain di luar variabel yang digunakan pada penelitian ini.

Setelah proses evaluasi selesai, model *Random Forest Regression* yang telah dibangun digunakan sebagai alat bantu prediksi permintaan layanan percetakan berdasarkan data historis transaksi. Tahap ini merupakan bentuk implementasi (*deployment*) model dalam penelitian, yaitu memanfaatkan model yang telah memperoleh hasil evaluasi untuk menghasilkan prediksi yang dapat digunakan sebagai dasar dalam perencanaan operasional dan pengambilan keputusan. Pada penelitian ini, *deployment* tidak dilakukan dalam bentuk aplikasi atau sistem informasi, melainkan sebagai implementasi model prediksi yang telah divalidasi menggunakan metode *TimeSeriesSplit*.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Random Forest Regression* dapat diterapkan untuk memprediksi permintaan layanan percetakan berdasarkan data historis transaksi pada ADIA SOLUTION dengan menggunakan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM). Proses *feature engineering*, *optimasi hyperparameter*, dan evaluasi menggunakan metode *TimeSeriesSplit* sebanyak lima *fold* menghasilkan nilai rata-rata *Mean Absolute Error* (MAE) sebesar 39,20, *Root Mean Squared Error* (RMSE) sebesar 63,62 dan *R-Squared* (R^2) sebesar 0,7071. Hasil tersebut menunjukkan bahwa model telah mampu mempelajari pola permintaan berdasarkan data historis transaksi dan dapat digunakan sebagai salah satu alternatif dalam membantu perencanaan operasional serta pengambilan keputusan pada layanan percetakan, meskipun masih terdapat variasi permintaan yang belum dapat dijelaskan secara optimal oleh model.

5. Saran

Penelitian selanjutnya disarankan menggunakan dataset dengan periode yang lebih panjang serta menambahkan variabel yang berpotensi memengaruhi permintaan layanan percetakan, seperti hari libur nasional, kalender akademik, promosi, dan faktor musiman, sehingga kemampuan model dalam menjelaskan variasi data dapat ditingkatkan. Selain itu, penelitian berikutnya dapat membandingkan performa *Random Forest Regression* dengan algoritma prediksi lainnya, seperti *XGBoost Regression*, *LightGBM*, atau *Gradient Boosting Regression*, untuk memperoleh model prediksi yang memberikan kinerja lebih baik sesuai dengan karakteristik data historis transaksi.

6. Daftar Pustaka

- [1] A. R. Hapriyanto, "Strategi inovatif dalam meningkatkan daya saing bisnis di era digital," *Nusantara Journal of Multidisciplinary Science*, vol. 2, no. 1, pp. 115–124, 2024.
- [2] V. Rahayu, B. Soedijo, and S. Sudarmawan, "Analisis Algoritma Apriori dan FP-Growth Dalam Menemukan Pola Frequent Item Data Association Rule Pada Supermarket," *Explore*, vol. 11, no. 2, pp. 20–23, 2021.
- [3] E. Purwanto, B. P. C. Utomo, H. Permatasari, and F. Mohd, "Comparative Analysis of Classification Models for Sales Prediction in E-commerce: Decision Tree, Random Forest, SVM, Naive Bayes, and KNN," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 6, pp. 5899–5915, 2025.
- [4] D. Yulianto and A. S. Nugraheni, "Efektivitas pembelajaran daring dalam pembelajaran bahasa Indonesia: Effectiveness of online learning in Indonesian language learning," *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 1, no. 1, pp. 33–42, 2021.
- [5] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS: Indonesia Journal Information System*, vol. 6, no. 2, pp. 83–89, 2023.
- [6] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Jurnal Sistem Informasi dan Sistem Komputer*, vol. 10, no. 1, pp. 101–109, 2025.
- [7] A. Julian, "Segmentasi pelanggan menggunakan algoritma K-Means pada jaringan telekomunikasi untuk optimalisasi strategi pemasaran," *JTT (Jurnal Teknologi Terpadu)*, vol. 13, no. 1, pp. 94–100, 2025.
- [8] A. P. Huslan, F. Mahardika, and D. I. Junaedi, "Implementasi Algoritma Random Forest untuk Prediksi Volume Pengunjung

- pada Sektor Layanan Publik (Studi Kasus Unit Layanan Publik di Jawa Barat Tahun 2021-2022),” *Riau Jurnal Teknik Informatika*, vol. 4, no. 3, pp. 446–â, 2025.
- [9] C. Sika Azkia, Y. Ansori, F. Darnis, and A. Rahman, “Pengaruh Feature Engineering Time Series Berbasis Rolling Statistics untuk Prediksi Suhu Menggunakan Random Forest dan Gradient Boosting,” vol. 11, no. 1, pp. 25–36, 2026, doi: 10.32493/jiup.v11i1.58137.
- [10] F. W. D. Wicaksono, M. Mula’ab, and A. B. Rahmat, “Optimasi Peramalan Tingkat Laju Inflasi Indonesia Melalui Pendekatan Ensemble Bagging Pada Algoritma Support Vector Regression,” *JTT (Jurnal Teknologi Terpadu)*, vol. 12, no. 2, pp. 144–150, 2024.
- [11] M. R. Maulizidan *et al.*, “Prediksi Gangguan Panik Menggunakan Knowledge Discovery in Database Dengan Algoritma Gradient Boosting,” *JTT (Jurnal Teknologi Terpadu)*, vol. 13, no. 2, pp. 166–174, 2025.