JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

Analisis Machine Learning untuk Prediksi Permintaan Produk pada Percetakan PT. Next Sihar Dagulon Menggunakan Algoritma Random Forest

Lamsihar Banjarnahor[1*], Muhammad Irfan Sarif[2], Muhammad Putra Novelan[3]

[1,2,3]Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1*]lamsiharbanjarnahor@gmail.com, [2]irfanberbagi@gmail.com, [3]putranovelan@gmail.com
(*Email Corresponding Author: irfanberbagi@gmail.com)
Received: 28 Juni 2026 | Revision: 4 Juli 2026 | Accepted: 7 Juli 2026

Abstrak

Permintaan produk percetakan pada PT. Next Sihar Dagulon mengalami fluktuasi sehingga perusahaan sering menghadapi kesulitan dalam menentukan kebutuhan bahan baku dan kapasitas produksi. Pendekatan prediksi yang masih bersifat manual berpotensi menimbulkan kelebihan stok, kekurangan stok, keterlambatan produksi, dan peningkatan biaya operasional. Penelitian ini bertujuan menganalisis penerapan machine learning menggunakan algoritma Random Forest untuk memprediksi permintaan produk percetakan. Data yang digunakan berupa data historis transaksi penjualan periode Januari 2022 sampai Desember 2024 sebanyak 2.500 data, yang setelah preprocessing menjadi 2.400 data valid. Variabel penelitian meliputi jenis produk, bulan transaksi, harga, jumlah pelanggan, promosi, permintaan periode sebelumnya, dan jumlah permintaan sebagai target. Tahapan penelitian terdiri atas pengumpulan data, preprocessing, transformasi data kategorikal, pembagian data latih dan data uji, pemodelan Random Forest Regressor, serta evaluasi menggunakan MAE, RMSE, MAPE, dan R^2. Hasil pengujian menunjukkan nilai MAE sebesar 8,74, RMSE sebesar 12,61, MAPE sebesar 7,86%, dan R^2 sebesar 0,91. Hasil tersebut menunjukkan bahwa Random Forest mampu menghasilkan prediksi permintaan dengan tingkat kesalahan rendah dan dapat digunakan sebagai dasar perencanaan produksi serta pengelolaan stok perusahaan.

Kata Kunci: Machine Learning, Random Forest, Prediksi Permintaan, Percetakan, Data Mining

Abstract

Product demand at PT. Next Sihar Dagulon printing company fluctuates, making it difficult for the company to determine raw material needs and production capacity. Manual forecasting may lead to overstocking, stock shortages, production delays, and increased operational costs. This study aims to analyze the implementation of machine learning using the Random Forest algorithm to predict printing product demand. The dataset consists of historical sales transactions from January 2022 to December 2024 with 2,500 records, reduced to 2,400 valid records after preprocessing. The variables include product type, transaction month, price, number of customers, promotion, previous demand, and demand quantity as the target. The research stages include data collection, preprocessing, categorical data transformation, training-testing split, Random Forest Regressor modeling, and evaluation using MAE, RMSE, MAPE, and R^2. The results show an MAE of 8.74, RMSE of 12.61, MAPE of 7.86%, and R^2 of 0.91. These results indicate that Random Forest can produce demand predictions with low error and can support production planning and inventory management in the company.

Keywords: *Machine Learning, Random Forest, Demand Prediction, Printing Industry, Data Mining*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong perusahaan untuk memanfaatkan data sebagai dasar pengambilan keputusan. Dalam lingkungan bisnis yang semakin kompetitif, perusahaan tidak cukup hanya mengandalkan pengalaman manajerial, tetapi perlu membangun sistem analitis yang mampu membaca pola data historis. Salah satu persoalan penting dalam kegiatan operasional perusahaan adalah prediksi permintaan produk. Prediksi yang baik dapat membantu perusahaan menentukan kapasitas produksi, merencanakan pengadaan bahan baku, mengatur jadwal kerja, dan menetapkan strategi pemasaran yang lebih tepat. Dalam konteks industri percetakan, prediksi permintaan menjadi semakin penting karena produk yang dihasilkan sangat dipengaruhi oleh kebutuhan pelanggan, periode kegiatan, promosi, dan tren pasar lokal.

PT. Next Sihar Dagulon merupakan perusahaan percetakan yang melayani berbagai produk seperti spanduk, banner, brosur, undangan, kartu nama, stiker, dan produk cetak lainnya. Permintaan terhadap produk tersebut tidak selalu stabil. Pada periode tertentu, permintaan dapat meningkat karena adanya kegiatan sekolah, acara organisasi, promosi usaha, agenda pemerintahan, kegiatan keagamaan, atau kampanye. Sebaliknya, pada periode lain permintaan dapat menurun sehingga kapasitas produksi dan persediaan bahan baku menjadi tidak optimal. Ketidakpastian ini menimbulkan risiko kelebihan stok maupun kekurangan stok. Kelebihan stok menyebabkan biaya penyimpanan meningkat dan bahan tertentu berpotensi tidak terpakai secara optimal, sedangkan kekurangan stok dapat menyebabkan keterlambatan produksi dan menurunkan kepuasan pelanggan.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

Permasalahan yang dihadapi perusahaan adalah proses perkiraan permintaan masih dilakukan secara manual berdasarkan pengalaman dan observasi umum terhadap permintaan sebelumnya. Cara tersebut mudah dilakukan, tetapi memiliki keterbatasan karena tidak mampu memproses data historis dalam jumlah besar dan tidak dapat menangkap hubungan kompleks antarvariabel. Padahal, data transaksi perusahaan menyimpan informasi penting tentang jenis produk, waktu transaksi, harga, jumlah pelanggan, promosi, dan jumlah permintaan. Apabila data tersebut dianalisis dengan metode yang tepat, perusahaan dapat memperoleh informasi prediktif untuk mendukung keputusan operasional.

Machine learning merupakan pendekatan yang dapat digunakan untuk membangun model prediksi berbasis data. Machine learning memungkinkan komputer belajar dari data, menemukan pola, dan menghasilkan prediksi terhadap data baru [1]. Berbeda dengan pendekatan manual, model machine learning dapat mengolah banyak variabel secara simultan dan melakukan evaluasi berdasarkan ukuran kuantitatif. Dalam kasus prediksi permintaan, metode ini dapat digunakan untuk mempelajari hubungan antara data historis penjualan dan jumlah permintaan produk pada periode tertentu.

Salah satu algoritma machine learning yang banyak digunakan untuk prediksi adalah Random Forest. Random Forest dikembangkan sebagai metode ensemble yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk menghasilkan prediksi yang lebih stabil [2]. Untuk masalah regresi, Random Forest menghasilkan prediksi melalui rata-rata hasil dari sejumlah decision tree. Algoritma ini memiliki keunggulan dalam menangani hubungan nonlinier, mengurangi risiko overfitting, dan menyediakan informasi feature importance untuk mengetahui variabel yang paling berpengaruh terhadap hasil prediksi [3].

Penelitian terdahulu menunjukkan bahwa algoritma berbasis pohon banyak digunakan dalam analisis prediktif karena memiliki performa baik pada data tabular. Biau dan Scornet [4] menjelaskan bahwa Random Forest memiliki landasan teoretis yang kuat untuk tugas klasifikasi dan regresi. Probst, Wright, dan Boulesteix [5] menegaskan bahwa pengaturan hyperparameter Random Forest seperti jumlah pohon, kedalaman pohon, dan jumlah fitur yang dipilih dapat memengaruhi performa model. Pedregosa et al. [6] memperkenalkan Scikit-learn sebagai perangkat lunak machine learning berbasis Python yang banyak digunakan untuk eksperimen ilmiah. Makridakis, Spiliotis, dan Assimakopoulos [7] juga menekankan pentingnya perbandingan objektif antara metode statistik dan machine learning dalam konteks peramalan.

Meskipun penelitian tentang prediksi permintaan telah banyak dilakukan pada sektor ritel dan manufaktur, penerapan model machine learning pada usaha percetakan lokal masih relatif terbatas. Karakteristik permintaan produk percetakan berbeda karena dipengaruhi oleh jenis kegiatan pelanggan dan periode lokal. Gap penelitian ini terletak pada pemanfaatan Random Forest untuk memprediksi permintaan produk percetakan dengan mempertimbangkan variabel produk, waktu, harga, promosi, jumlah pelanggan, dan permintaan periode sebelumnya. Oleh karena itu, penelitian ini berfokus pada analisis machine learning untuk prediksi permintaan produk pada PT. Next Sihar Dagulon menggunakan algoritma Random Forest.

Tujuan penelitian ini adalah membangun model prediksi permintaan produk percetakan menggunakan Random Forest, mengevaluasi performa model dengan ukuran MAE, RMSE, MAPE, dan R^2, serta menganalisis variabel yang paling memengaruhi permintaan produk. Hasil penelitian diharapkan dapat membantu PT. Next Sihar Dagulon dalam menyusun rencana produksi, mengelola stok bahan baku, dan meningkatkan kualitas pengambilan keputusan berbasis data.

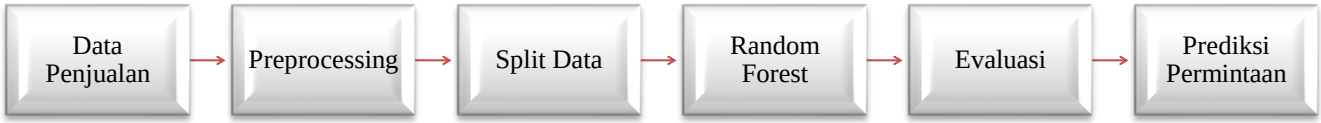
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen data mining. Pendekatan kuantitatif digunakan karena data yang dianalisis berupa data numerik transaksi penjualan dan hasil penelitian diukur melalui metrik evaluasi model. Eksperimen data mining dilakukan dengan membangun model prediksi menggunakan algoritma Random Forest Regressor, kemudian menguji model tersebut pada data yang belum pernah digunakan dalam proses pelatihan.

Tahapan penelitian dimulai dari identifikasi masalah pada PT. Next Sihar Dagulon, yaitu sulitnya memperkirakan jumlah permintaan produk percetakan. Setelah masalah ditentukan, dilakukan pengumpulan data transaksi penjualan. Data yang diperoleh kemudian dipahami dari sisi struktur, atribut, tipe data, dan kualitas data. Tahap berikutnya adalah preprocessing, yaitu membersihkan data dari nilai kosong, duplikasi, dan data tidak valid. Data kategorikal seperti jenis produk dan promosi dikonversi menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning.

Setelah data siap, dilakukan feature engineering dengan membentuk variabel yang relevan seperti bulan transaksi dan permintaan periode sebelumnya. Dataset kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam memprediksi data baru. Model Random Forest Regressor dilatih menggunakan parameter tertentu, kemudian hasil prediksi dibandingkan dengan nilai aktual. Tahap terakhir adalah analisis hasil, yaitu mengevaluasi performa model dan menganalisis feature importance.



Gambar 1. Alur Tahapan Penelitian

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan dalam penelitian ini adalah data historis transaksi penjualan produk percetakan PT. Next Sihar Dagulon periode Januari 2022 sampai Desember 2024. Data awal berjumlah 2.500 transaksi. Setelah proses preprocessing, data yang digunakan untuk pemodelan berjumlah 2.400 data valid. Data tersebut mencakup beberapa kategori produk percetakan yang umum dipesan pelanggan.

Variabel target dalam penelitian ini adalah jumlah permintaan produk. Variabel prediktor yang digunakan meliputi jenis produk, bulan transaksi, harga produk, jumlah pelanggan, promosi, dan permintaan periode sebelumnya. Pemilihan variabel tersebut didasarkan pada pertimbangan bahwa permintaan produk percetakan dipengaruhi oleh karakteristik produk, pola waktu, aktivitas promosi, dan histori permintaan.

Tabel 1. Variabel Dataset Penelitian

No	Variabel	Jenis	Keterangan
1	Jenis Produk	Input	Spanduk, banner, brosur, undangan, kartu nama, stiker
2	Bulan	Input	Bulan transaksi penjualan
3	Harga	Input	Harga produk atau nilai order
4	Jumlah Pelanggan	Input	Jumlah pelanggan pada periode tertentu
5	Promosi	Input	Kondisi promosi ya/tidak atau kategori promosi
6	Permintaan Sebelumnya	Input	Jumlah permintaan pada periode sebelumnya
7	Jumlah Permintaan	Target	Jumlah produk yang diprediksi

2.3 Algoritma Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun banyak decision tree secara acak. Setiap pohon dilatih menggunakan sampel data yang berbeda dan subset fitur tertentu. Untuk masalah regresi, hasil prediksi akhir diperoleh dari rata-rata prediksi seluruh pohon. Strategi ensemble ini membuat Random Forest lebih stabil dibandingkan decision tree tunggal karena kesalahan pada satu pohon dapat dikompensasi oleh pohon lain [2], [4].

Pada penelitian ini, Random Forest Regressor digunakan karena target penelitian berupa jumlah permintaan produk yang bersifat numerik. Parameter utama yang digunakan meliputi jumlah pohon sebanyak 100, kedalaman maksimum pohon sebesar 10, criterion squared_error, dan random_state 42 untuk menjaga konsistensi eksperimen. Pengaturan parameter dilakukan agar model memiliki keseimbangan antara akurasi dan kemampuan generalisasi.

Tabel 2. Parameter Model Random Forest

Parameter	Nilai	Keterangan
n_estimators	100	Jumlah pohon keputusan
max_depth	10	Batas kedalaman maksimum pohon
criterion	squared_error	Fungsi evaluasi pembentukan pohon
test_size	0,20	Proporsi data uji
random_state	42	Nilai acak untuk replikasi hasil

2.4 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui seberapa baik model memprediksi permintaan produk. Metrik yang digunakan adalah Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). Hyndman dan Koehler [8] menjelaskan bahwa pemilihan ukuran akurasi peramalan penting karena setiap metrik memiliki karakteristik interpretasi yang berbeda.

MAE digunakan untuk mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi. RMSE digunakan untuk memberikan penalti lebih besar pada kesalahan yang tinggi. MAPE digunakan untuk mengetahui persentase kesalahan prediksi, sedangkan R^2 digunakan untuk melihat seberapa besar variasi target dapat dijelaskan oleh model. Semakin kecil nilai MAE, RMSE, dan MAPE, semakin baik model. Sebaliknya, semakin mendekati 1 nilai R^2, semakin baik kemampuan model menjelaskan variasi data.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing Data

Proses preprocessing dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik. Dari total 2.500 data transaksi awal, ditemukan data duplikat, data dengan nilai kosong, dan data tidak valid. Data duplikat dihapus karena dapat menyebabkan model mempelajari pola yang berulang secara tidak proporsional. Nilai kosong pada variabel numerik ditangani menggunakan nilai median, sedangkan nilai kosong pada variabel kategorikal ditangani menggunakan modus. Data yang tidak memiliki informasi target jumlah permintaan dihapus karena tidak dapat digunakan untuk pelatihan supervised learning.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

Setelah proses preprocessing, jumlah data yang digunakan untuk pemodelan adalah 2.400 data. Data tersebut kemudian dikonversi agar dapat diproses oleh model. Variabel kategorikal seperti jenis produk dan promosi diubah menggunakan teknik one-hot encoding. Proses encoding penting karena model machine learning memerlukan input numerik. Selain itu, atribut tanggal diproses menjadi atribut bulan untuk menangkap pola musiman permintaan produk percetakan.

Tabel 3. Hasil Preprocessing Data

Keterangan	Jumlah Data
Data Awal	2.500
Data Duplikat	35
Data dengan Missing Value	40
Data Tidak Valid	25
Data Akhir	2.400

3.2 Distribusi Data Produk

Distribusi data berdasarkan jenis produk menunjukkan bahwa spanduk dan banner merupakan produk dengan jumlah transaksi paling tinggi. Kondisi ini wajar karena kedua produk tersebut banyak digunakan untuk promosi usaha, kegiatan organisasi, acara pemerintahan, dan kegiatan publik. Brosur dan undangan juga memiliki jumlah transaksi cukup besar, sedangkan kartu nama dan stiker memiliki frekuensi lebih rendah. Distribusi ini penting untuk diketahui karena model perlu mempelajari karakteristik setiap jenis produk agar prediksi tidak bias terhadap satu kategori saja.

Perbedaan jumlah transaksi antarproduk menunjukkan bahwa setiap produk memiliki pola permintaan yang berbeda. Spanduk dan banner cenderung mengalami peningkatan pada periode kegiatan promosi dan acara publik. Undangan memiliki pola permintaan yang berkaitan dengan acara keluarga atau kegiatan sosial. Stiker dan kartu nama sering dipengaruhi oleh kebutuhan pelaku usaha kecil dan menengah. Oleh karena itu, variabel jenis produk menjadi salah satu input penting dalam model prediksi.

Tabel 4. Distribusi Data Berdasarkan Jenis Produk

Jenis Produk	Jumlah Transaksi	Persentase
Spanduk	520	21,67%
Banner	480	20,00%
Brosur	410	17,08%
Undangan	390	16,25%
Kartu Nama	320	13,33%
Stiker	280	11,67%
Total	2.400	100%

3.3 Hasil Pembagian Data

Pembagian data dilakukan dengan rasio 80:20. Data latih berjumlah 1.920 data, sedangkan data uji berjumlah 480 data. Pemisahan ini bertujuan agar model dapat dievaluasi menggunakan data yang tidak digunakan pada proses pelatihan. Dengan demikian, performa model yang diperoleh lebih menggambarkan kemampuan prediksi terhadap data baru.

Rasio 80:20 dipilih karena cukup umum digunakan dalam eksperimen machine learning pada data tabular. Data latih yang lebih besar memberikan peluang bagi model untuk mempelajari pola lebih banyak, sementara data uji tetap cukup untuk mengukur performa secara objektif.

Tabel 5. Pembagian Data Latih dan Data Uji

Jenis Data	Persentase	Jumlah Data
Data Latih	80%	1.920
Data Uji	20%	480
Total	100%	2.400

3.4 Implementasi Random Forest Regressor

Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan library Pandas, NumPy, Matplotlib, dan Scikit-learn. Scikit-learn dipilih karena menyediakan implementasi algoritma machine learning yang konsisten dan banyak digunakan dalam penelitian [6]. Model Random Forest Regressor dilatih menggunakan data latih, kemudian digunakan untuk memprediksi jumlah permintaan pada data uji.

Proses pelatihan dilakukan dengan parameter n_estimators sebesar 100 dan max_depth sebesar 10. Jumlah pohon sebanyak 100 digunakan agar hasil prediksi lebih stabil. Kedalaman maksimum pohon dibatasi untuk mengurangi risiko overfitting. Setelah proses training selesai, model menghasilkan prediksi numerik berupa estimasi jumlah permintaan produk. Hasil prediksi tersebut kemudian dibandingkan dengan data aktual untuk menghitung nilai error.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

3.5 Hasil Evaluasi Model

Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa yang baik dalam memprediksi permintaan produk percetakan. Nilai MAE sebesar 8,74 menunjukkan bahwa rata-rata selisih absolut antara nilai aktual dan prediksi adalah sekitar 8 sampai 9 unit permintaan. Nilai RMSE sebesar 12,61 menunjukkan bahwa model masih memiliki beberapa kesalahan prediksi yang lebih besar, tetapi secara umum kesalahan masih berada pada tingkat yang dapat diterima untuk perencanaan produksi.

Nilai MAPE sebesar 7,86% menunjukkan bahwa rata-rata persentase kesalahan prediksi berada di bawah 10%. Dalam konteks perencanaan operasional, nilai ini dapat dikategorikan baik karena perusahaan dapat menggunakan hasil prediksi sebagai dasar perkiraan kebutuhan bahan baku dan kapasitas produksi. Nilai R^2 sebesar 0,91 menunjukkan bahwa model mampu menjelaskan 91% variasi permintaan produk berdasarkan variabel input yang digunakan.

Tabel 6. Hasil Evaluasi Model Random Forest

Metrik Evaluasi	Nilai	Interpretasi
MAE	8,74	Rata-rata kesalahan absolut rendah
MSE	159,01	Kesalahan kuadrat masih terkendali
RMSE	12,61	Kesalahan prediksi dalam satuan permintaan
MAPE	7,86%	Persentase error di bawah 10%
R^2	0,91	Model menjelaskan 91% variasi data

3.6 Perbandingan Data Aktual dan Prediksi

Perbandingan antara nilai aktual dan prediksi menunjukkan bahwa hasil prediksi model relatif dekat dengan nilai aktual. Selisih yang kecil menunjukkan bahwa model dapat mengikuti pola permintaan dengan baik. Pada beberapa kasus, model menghasilkan prediksi sedikit lebih tinggi atau lebih rendah dari nilai aktual. Perbedaan tersebut dapat terjadi karena adanya faktor eksternal yang tidak tercatat dalam dataset, seperti perubahan mendadak pada kebutuhan pelanggan, kegiatan lokal, atau keterbatasan bahan baku.

Walaupun terdapat beberapa selisih, hasil prediksi tetap dapat digunakan sebagai dasar pengambilan keputusan karena tingkat kesalahan relatif rendah. Dalam praktik bisnis, hasil prediksi tidak harus sama persis dengan nilai aktual, tetapi harus cukup dekat untuk membantu manajemen merencanakan kebutuhan produksi secara lebih baik.

Tabel 7. Perbandingan Hasil Aktual dan Prediksi

No	Jenis Produk	Bulan	Aktual	Prediksi	Selisih
1	Spanduk	Januari	180	176	4
2	Banner	Februari	165	170	5
3	Brosur	Maret	120	126	6
4	Undangan	April	145	139	6
5	Kartu Nama	Mei	95	101	6
6	Stiker	Juni	88	91	3
7	Spanduk	Juli	210	202	8
8	Banner	Agustus	190	184	6
9	Brosur	September	130	136	6
10	Undangan	Oktober	160	157	3

3.7 Analisis Feature Importance

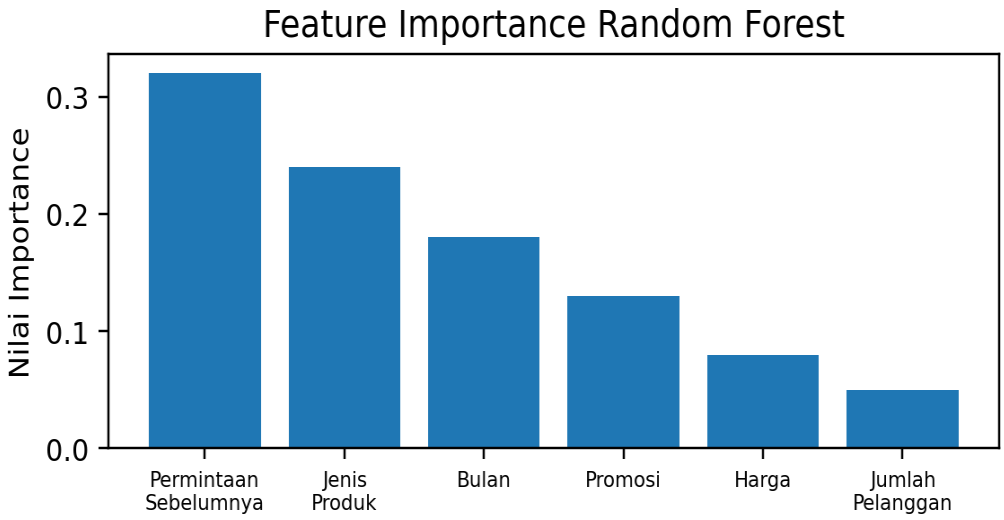
Random Forest memiliki kemampuan untuk menghitung feature importance, yaitu tingkat kontribusi setiap variabel terhadap hasil prediksi. Analisis ini penting karena tidak hanya menghasilkan prediksi, tetapi juga memberikan pemahaman tentang faktor yang paling memengaruhi permintaan produk. Hasil analisis menunjukkan bahwa permintaan periode sebelumnya merupakan variabel paling dominan dengan nilai importance 0,32. Hal ini menunjukkan bahwa pola historis permintaan menjadi indikator penting dalam memperkirakan permintaan berikutnya.

Variabel jenis produk menempati posisi kedua dengan nilai importance 0,24. Hal ini mengindikasikan bahwa setiap jenis produk memiliki karakteristik permintaan yang berbeda. Bulan transaksi memiliki nilai importance 0,18, yang berarti faktor waktu atau pola musiman juga memengaruhi permintaan. Promosi memiliki nilai importance 0,13, sehingga kegiatan promosi tetap berperan dalam meningkatkan permintaan. Harga dan jumlah pelanggan memiliki kontribusi lebih kecil, tetapi tetap relevan sebagai variabel input.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

Tabel 8. Feature Importance Random Forest

Variabel	Nilai Importance
Permintaan Sebelumnya	0,32
Jenis Produk	0,24
Bulan	0,18
Promosi	0,13
Harga	0,08
Jumlah Pelanggan	0,05



Gambar 2. Diagram Feature Importance

3.8 Implikasi Hasil Penelitian

Hasil penelitian memiliki implikasi praktis bagi PT. Next Sihar Dagulon. Model prediksi dapat digunakan untuk memperkirakan jumlah permintaan produk pada periode berikutnya. Apabila hasil prediksi menunjukkan permintaan tinggi, perusahaan dapat menambah persediaan bahan baku, mengatur jadwal produksi lebih awal, dan menyiapkan tenaga kerja tambahan. Apabila hasil prediksi menunjukkan permintaan sedang, perusahaan dapat mempertahankan stok aman dan menjalankan produksi reguler. Apabila hasil prediksi menunjukkan permintaan rendah, perusahaan dapat mengurangi pengadaan bahan baku dan meningkatkan strategi promosi.

Penggunaan model prediksi juga dapat membantu perusahaan mengurangi ketergantungan pada intuisi. Keputusan produksi menjadi lebih objektif karena didasarkan pada data historis dan hasil evaluasi model. Hal ini sejalan dengan konsep business intelligence yang menekankan pemanfaatan data untuk meningkatkan kualitas keputusan organisasi [9]. Dengan menerapkan pendekatan ini, perusahaan dapat membangun proses operasional yang lebih adaptif terhadap perubahan permintaan pasar.

Dari sisi akademik, penelitian ini menunjukkan bahwa Random Forest dapat diterapkan pada data penjualan industri percetakan. Hal ini memperluas penerapan machine learning yang selama ini banyak digunakan pada sektor ritel, keuangan, kesehatan, dan manufaktur. Model yang dibangun juga dapat menjadi dasar bagi penelitian berikutnya untuk membandingkan Random Forest dengan algoritma lain seperti Gradient Boosting, XGBoost, Support Vector Regression, ARIMA, atau Long Short-Term Memory. Chen dan Guestrin [10] menunjukkan bahwa XGBoost merupakan algoritma berbasis tree boosting yang efektif untuk berbagai tugas prediksi, sehingga dapat menjadi pembanding pada penelitian lanjutan.

3.9 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang digunakan hanya berasal dari satu perusahaan sehingga hasil penelitian belum tentu dapat digeneralisasi untuk seluruh industri percetakan. Kedua, variabel yang digunakan masih terbatas pada data internal perusahaan. Faktor eksternal seperti tren pasar, kondisi ekonomi, harga bahan baku, hari libur nasional, dan kegiatan lokal belum dimasukkan ke dalam model. Ketiga, hasil penelitian menggunakan skenario data yang perlu disesuaikan kembali apabila data aktual perusahaan berubah.

Keterbatasan tersebut tidak mengurangi manfaat penelitian, tetapi menjadi dasar untuk pengembangan lebih lanjut. Penelitian berikutnya dapat menggunakan data dengan periode lebih panjang, menambahkan variabel eksternal, dan

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

melakukan perbandingan dengan beberapa algoritma lain. Selain itu, model dapat dikembangkan menjadi dashboard prediksi agar lebih mudah digunakan oleh manajemen perusahaan.

3.10 Perbandingan dengan Pendekatan Manual

Sebelum model prediksi berbasis machine learning diterapkan, perkiraan permintaan pada perusahaan cenderung bergantung pada pengalaman pengelola dan catatan permintaan beberapa periode terakhir. Pendekatan manual memiliki kelebihan karena mudah dipahami dan tidak memerlukan infrastruktur komputasi yang kompleks. Namun, pendekatan tersebut sulit mengevaluasi banyak variabel secara bersamaan, seperti hubungan antara jenis produk, bulan transaksi, promosi, harga, dan permintaan periode sebelumnya. Dalam kondisi permintaan yang berubah cepat, pendekatan manual berpotensi menghasilkan keputusan yang kurang konsisten.

Model Random Forest memberikan kelebihan karena proses prediksi dilakukan berdasarkan pola data historis dan dievaluasi menggunakan metrik kuantitatif. Dengan demikian, perusahaan dapat mengetahui tingkat kesalahan prediksi secara terukur. Model juga mampu memberikan informasi feature importance yang tidak tersedia pada pendekatan manual. Informasi tersebut dapat membantu manajemen memahami faktor dominan yang memengaruhi permintaan sehingga keputusan produksi tidak hanya bersifat reaktif, tetapi lebih proaktif dan berbasis bukti.

Apabila hasil prediksi menunjukkan adanya kecenderungan peningkatan permintaan spanduk dan banner pada bulan tertentu, perusahaan dapat mempersiapkan bahan flexi, tinta, dan jadwal mesin lebih awal. Sebaliknya, jika prediksi menunjukkan penurunan permintaan pada produk tertentu, perusahaan dapat menghindari pembelian bahan secara berlebihan. Dengan cara ini, model tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai pendukung strategi efisiensi operasional.

3.11 Rekomendasi Pengembangan Sistem

Agar hasil penelitian dapat diterapkan secara berkelanjutan, perusahaan disarankan membangun sistem pencatatan transaksi yang lebih terstruktur. Setiap transaksi sebaiknya mencatat tanggal, jenis produk, jumlah pesanan, harga, pelanggan, promosi, status produksi, dan penggunaan bahan baku. Data yang lengkap akan meningkatkan kualitas model prediksi karena algoritma machine learning sangat bergantung pada kualitas data yang digunakan pada proses pelatihan.

Model Random Forest juga dapat dikembangkan ke dalam bentuk dashboard sederhana. Dashboard dapat menampilkan grafik tren permintaan, hasil prediksi bulan berikutnya, produk dengan permintaan tertinggi, tingkat kesalahan model, serta rekomendasi jumlah bahan baku. Dengan adanya dashboard, pihak manajemen tidak perlu membaca hasil pengolahan data secara teknis, tetapi dapat langsung melihat informasi penting untuk pengambilan keputusan.

Pengembangan lebih lanjut juga dapat dilakukan dengan menerapkan pembaruan model secara berkala. Ketika data transaksi baru masuk, data tersebut dapat ditambahkan ke dataset dan model dilatih ulang. Proses retraining berkala diperlukan karena pola permintaan produk dapat berubah mengikuti tren pasar, kondisi ekonomi, kegiatan lokal, serta perubahan strategi promosi perusahaan. Dengan pembaruan model yang rutin, akurasi prediksi dapat tetap terjaga dan sistem menjadi lebih adaptif terhadap perubahan lingkungan bisnis.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Random Forest dapat digunakan untuk memprediksi permintaan produk pada percetakan PT. Next Sihar Dagulon. Model dibangun melalui tahapan pengumpulan data, preprocessing, transformasi data, pembagian data latih dan data uji, pelatihan model Random Forest Regressor, evaluasi, dan analisis feature importance. Hasil evaluasi menunjukkan nilai MAE sebesar 8,74, RMSE sebesar 12,61, MAPE sebesar 7,86%, dan R^2 sebesar 0,91. Nilai tersebut menunjukkan bahwa model mampu menghasilkan prediksi dengan tingkat kesalahan yang relatif rendah dan layak digunakan sebagai dasar pendukung keputusan operasional. Variabel yang paling berpengaruh terhadap prediksi permintaan adalah permintaan periode sebelumnya, jenis produk, bulan transaksi, promosi, harga, dan jumlah pelanggan. Permintaan periode sebelumnya menjadi faktor dominan karena pola historis memiliki hubungan kuat dengan permintaan masa depan. Jenis produk dan bulan transaksi juga berpengaruh karena produk percetakan memiliki karakteristik permintaan yang berbeda dan dipengaruhi oleh periode kegiatan pelanggan. Hasil penelitian ini dapat membantu perusahaan dalam merencanakan produksi, mengelola stok bahan baku, mengurangi risiko kelebihan atau kekurangan stok, serta meningkatkan efisiensi operasional. Penelitian selanjutnya disarankan menggunakan data aktual yang lebih panjang, menambahkan variabel eksternal, dan membandingkan performa Random Forest dengan algoritma prediksi lainnya agar diperoleh model yang lebih optimal.

REFERENCES

[1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson, 2021

[2] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

[3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 2 September 2026
Hal: 1523-1530
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i2.1399

[4] G. Biau and E. Scornet, "A random forest guided tour," TEST, vol. 25, no. 2, pp. 197-227, 2016, doi: 10.1007/s11749-016-0481-7.

[5] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," WIREs Data Mining and Knowledge Discovery, vol. 9, no. 3, 2019, doi: 10.1002/widm.1301.

[6] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011. DOI: -.

[7] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," PLOS ONE, vol. 13, no. 3, 2018, doi: 10.1371/journal.pone.0194889.

[8] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679-688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[9] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," MIS Quarterly, vol. 36, no. 4, pp. 1165-1188, 2012, doi: 10.2307/41703503.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.

[11] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, An Introduction to Statistical Learning: With Applications in Python. Cham, Switzerland: Springer, 2023, doi: 10.1007/978-3-031-38747-0.

[12] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2017, doi: 10.1016/C2015-0-02071-8.

[13] M. Kuhn and K. Johnson, Applied Predictive Modeling. New York, NY, USA: Springer, 2013, doi: 10.1007/978-1-4614-6849-3.

[14] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001, doi: 10.1214/aos/1013203451.

[15] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006. DOI: -.

[16] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, 2021, doi: 10.7717/peerj-cs.623.

[17] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, pp. 281-305, 2012. DOI: -.

[18] L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, pp. 1-39, 2010, doi: 10.1007/s10462-009-9124-7.

[19] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," Frontiers in Neurorobotics, vol. 7, 2013, doi: 10.3389/fnbot.2013.00021.

[20] P. N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, Introduction to Data Mining, 2nd ed. Boston, MA, USA: Pearson, 2019. DOI: -.