

Predicting Consumer Demand At Nayla Buah Store, Ratu Agung District, Bengkulu City Using Machine Learning Technology

Yulia Darnita ¹⁾; Muntahanah ²⁾; Dandi Sunardi ³⁾; Ahmad Fauzi ⁴⁾

^{1,2,3,4)} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu

Email: ¹⁾ yuliadarnita@umb.ac.id

How to Cite :

Darnita, Y., Muntahanah, Sunardi, D., Fauzi, A. (2026). Predicting Consumer Demand at Nayla Buah Store, Ratu Agung District, Bengkulu City Using Machine Learning Technology. Jurnal Media Computer Science, 5 (4).

ARTICLE HISTORY

Received [02 September 2026]

Revised [23 Oktober 2026]

Accepted [25 Oktober 2026]

KEYWORDS

Prediction, Demand, Random Forest Regressor, K-Nearest Neighbor (KNN).

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Industri ritel, termasuk toko buah, sangat dipengaruhi oleh dinamika permintaan konsumen yang fluktuatif. Ketidakmampuan untuk memprediksi permintaan secara akurat dapat menimbulkan berbagai masalah operasional bagi toko buah yaitu kelebihan stok (*overstock*) dan kekurangan stok (*understock*). Oleh karena itu, kemampuan untuk memperkirakan permintaan secara akurat menjadi kebutuhan penting bagi pelaku usaha. Metode *Random Forest Regressor* dan *K-Nearest Neighbor* (KNN) memberikan gambaran evaluatif mengenai algoritma mana yang paling sesuai digunakan untuk memprediksi permintaan. Hasil penelitian menunjukkan bahwa faktor jenis buah, faktor hari, promosi, dan cuaca memiliki pengaruh terhadap jumlah permintaan buah oleh konsumen. Variabel-variabel tersebut mampu digunakan sebagai dasar dalam melakukan prediksi permintaan buah pada toko buah. Hasil prediksi dari kedua model dapat digunakan sebagai dasar perencanaan stok dan pengambilan keputusan operasional. Toko dapat mengurangi risiko kerugian akibat kelebihan stok maupun kekurangan stok dengan memperkirakan permintaan harian secara lebih akurat. Secara keseluruhan, *Random Forest Regressor* adalah model yang paling direkomendasikan untuk diterapkan dalam memprediksi permintaan buah di Toko Nayla Buah karena memiliki tingkat akurasi yang lebih tinggi, stabil, dan mampu memproses variasi data secara lebih efektif.

ABSTRACT

The retail industry, including fruit stores, is significantly influenced by fluctuating consumer demand dynamics. An inability to accurately predict demand can lead to various operational issues for fruit stores, namely overstocking and understocking. Therefore, the ability to accurately forecast demand is crucial for business owners. The *Random Forest Regressor* and *K-Nearest Neighbor* (KNN) methods provide an evaluation of which algorithm is most suitable for predicting demand. Research results indicate that factors such as fruit type, day of the week, promotions, and weather influence the volume of consumer demand for fruit. These variables can serve as a basis for predicting fruit demand at the store. The prediction results from both models can be used to inform stock planning and operational decision-making. By forecasting daily demand more accurately, the store can mitigate the risk of losses associated with both overstocking and understocking. Overall, the *Random Forest Regressor* is the most recommended model for predicting fruit demand at Nayla Buah Store, as it offers higher accuracy and stability, and is capable of processing data variations more effectively.

PENDAHULUAN

Industri ritel, termasuk toko buah, sangat dipengaruhi oleh dinamika permintaan konsumen yang fluktuatif. Ketidakmampuan untuk memprediksi permintaan secara akurat dapat menimbulkan berbagai masalah operasional bagi toko buah yaitu kelebihan stok (*overstock*) dan kekurangan stok (*understock*). Oleh karena itu, kemampuan untuk memperkirakan permintaan secara akurat menjadi kebutuhan penting bagi pelaku usaha.

Di era digital saat ini, data penjualan yang dikumpulkan setiap hari dapat dimanfaatkan sebagai sumber informasi yang sangat berharga untuk dianalisis menggunakan teknologi *machine learning* (pembelajaran mesin). Teknologi ini mampu mempelajari pola dari data historis dan membuat prediksi terhadap nilai di masa depan dengan tingkat akurasi yang tinggi. Selain itu, penerapan teknologi *machine learning* dalam usaha kecil seperti toko buah menunjukkan bagaimana transformasi digital tidak hanya relevan bagi perusahaan besar, tetapi juga dapat dimanfaatkan oleh pelaku usaha mikro untuk meningkatkan efisiensi dan daya saing. Dengan memanfaatkan data penjualan harian dan menganalisisnya menggunakan model yang tepat, pelaku usaha dapat mengambil keputusan berbasis data (*data driven decision making*) yang lebih rasional dan terukur dibandingkan hanya mengandalkan intuisi semata. Salah satu metode *machine learning* yang banyak digunakan dalam prediksi adalah *Random Forest Regressor*. Metode ini termasuk ke dalam kelompok *ensemble learning*, yaitu pendekatan yang menggabungkan banyak pohon keputusan (*decision tree*) untuk menghasilkan prediksi yang lebih stabil dan akurat. Keunggulan dari metode ini adalah kemampuannya dalam menangani data dengan variabel yang kompleks dan bersifat non-linear, serta relatif tahan terhadap *overfitting* dibandingkan model regresi tunggal. *Random Forest Regressor* dianggap sangat sesuai untuk digunakan dalam memprediksi permintaan konsumen, di mana variabel-variabelnya bersifat beragam dan tidak selalu memiliki hubungan yang linier. Di sisi lain, metode *K-Nearest Neighbor* (KNN) dipilih sebagai pembanding karena kesederhanaannya, fleksibilitasnya, serta kemampuannya dalam memberikan prediksi yang baik pada data yang memiliki pola kedekatan antar titik data. KNN bekerja berdasarkan prinsip bahwa data dengan karakteristik yang mirip cenderung memiliki nilai output yang mirip pula. Metode ini sangat efektif digunakan ketika data penjualan memiliki pola musiman atau kemiripan dari periode tertentu ke periode lainnya. Selain itu, KNN tidak memerlukan asumsi distribusi data, sehingga dapat langsung bekerja pada data apa adanya setelah dilakukan normalisasi.

Kombinasi kedua metode ini memberikan gambaran evaluatif mengenai algoritma mana yang paling sesuai digunakan untuk memprediksi permintaan di Toko Nayla Buah. *Random forest* mampu menganalisis pola kompleks secara global, sementara KNN mampu menangkap pola kedekatan secara lokal. Perbandingan ini diharapkan menghasilkan model prediksi terbaik yang dapat membantu pemilik usaha dalam menentukan jumlah stok yang optimal, meminimalkan risiko kelebihan atau kekurangan persediaan, serta meningkatkan efisiensi operasional toko.

Toko Nayla Buah merupakan salah satu usaha mikro yang menyediakan berbagai jenis buah segar seperti apel, jeruk, semangka, dan pisang. Dalam kegiatan operasionalnya, pemilik toko sering menghadapi tantangan dalam menentukan jumlah stok yang ideal setiap harinya sehingga diperlukan kemampuan prediksi permintaan yang tepat dan akurat. Ketidaktepatan dalam memperkirakan jumlah permintaan seringkali menyebabkan kerugian karena sebagian buah menjadi tidak layak jual akibat pembusukan, atau sebaliknya, pelanggan tidak terlayani karena stok sudah habis.

LANDASAN TEORI

Kecerdasan Buatan (*Artificial Intelligence*)

Artificial Intelligence (AI) merupakan bidang ilmu komputer yang bertujuan untuk menciptakan mesin atau program komputer yang dapat melakukan tugas yang memerlukan kecerdasan manusia, seperti pengenalan pola, pengambilan keputusan, dan pemrosesan bahasa alami (Juditha, 2025). AI telah diterapkan secara luas pada berbagai bidang, mulai dari peramalan cuaca, sistem penerjemahan, sistem rekomendasi, hingga pembuatan konten digital.

Deepfake

Deepfake merupakan media sintetis yang dihasilkan menggunakan teknologi AI untuk menampilkan sesuatu yang sebenarnya tidak pernah terjadi. Istilah *Deepfake* berasal dari gabungan kata *Deep learning*, yang merujuk pada teknologi pembelajaran mesin yang mendalam, dan *fake* yang berarti palsu. Kemunculan *Deepfake* pertama kali digunakan untuk memberikan kemudahan bagi manusia dalam menciptakan hasil foto atau video yang lebih nyata dalam konten hiburan seperti film atau gim, namun seiring berjalannya waktu teknologi ini mulai disalah gunakan untuk berbagai kepentingan negatif (Seveney, Wicaksono, & Soetijono, 2025).

Generative Adversarial Network (GAN)

GAN merupakan metode utama yang digunakan dalam pembuatan konten *Deepfake*, yang bekerja dengan menggunakan dua model jaringan saraf tiruan yang saling bersaing, yaitu generator yang menghasilkan konten palsu dan discriminator yang berupaya membedakan konten palsu dengan konten asli. Proses ini berlangsung berulang kali sehingga menghasilkan konten yang semakin realistis dan sulit dibedakan dari konten asli (Altuncu, Franqueira, & Li, 2024).

Disinformasi dan Hoaks Digital

Disinformasi merupakan informasi yang sengaja dibuat salah dengan tujuan menyesatkan atau memanipulasi opini publik. Media sosial menjadi saluran utama penyebaran hoaks karena keterbukaan platform dalam pertukaran informasi tanpa pengawasan dan penyaringan yang ketat (Sabela, Saputra, & Kuncoro, 2026).

Literasi Digital

Literasi digital merupakan kemampuan seseorang dalam memahami, mengevaluasi, dan menggunakan informasi digital secara kritis. Tingkat literasi digital menjadi faktor penentu kemampuan seseorang dalam mengenali dan menghindari dampak negatif dari konten manipulatif, termasuk *Deepfake* (Payne, 2026).

Metode Deskriptif Kuantitatif

Metode penelitian kuantitatif merupakan pendekatan ilmiah yang bertujuan untuk mengukur, menjelaskan, dan menguji hubungan antar variabel melalui data numerik yang dianalisis secara statistik, sehingga memungkinkan peneliti menarik kesimpulan yang objektif, teruji, dan dapat di generalisasikan ke populasi yang lebih luas (Hair et al., 2022). Metode deskriptif kuantitatif secara khusus bertujuan untuk menggambarkan dan menganalisis fenomena secara sistematis berdasarkan data numerik yang dikumpulkan dari lapangan.

Kuesioner sebagai Instrumen Penelitian

Kuesioner merupakan instrumen pengumpulan data yang berisi serangkaian pertanyaan atau pernyataan yang disusun secara sistematis untuk memperoleh informasi dari responden. Dalam penelitian ini, kuesioner disusun dengan menyertakan materi stimulus berupa pasangan konten (asli dan *Deepfake*) yang harus dinilai oleh responden, disertai pertanyaan mengenai pengetahuan dan pengalaman responden terkait teknologi *Deepfake* (Mamuaya et al., 2025).

Pengetahuan tentang Teknologi Deepfake

Pengetahuan tentang teknologi *Deepfake* mengacu pada sejauh mana seseorang memahami konsep, cara kerja, serta ciri-ciri konten yang dihasilkan oleh teknologi *Deepfake*. Pemahaman ini mencakup kesadaran akan keberadaan teknologi *Deepfake*, pengetahuan bahwa konten tersebut dihasilkan melalui algoritma AI seperti *Generative Adversarial Network (GAN)*, serta kemampuan mengenali indikator visual maupun audio yang membedakan konten asli dari konten manipulasi (Payne, 2026; Juditha, 2025).

Semakin tinggi tingkat pengetahuan seseorang mengenai karakteristik dan mekanisme kerja *Deepfake*, semakin besar pula potensi mereka untuk lebih waspada dan kritis ketika mengonsumsi konten digital. Dalam penelitian ini, pengetahuan tentang *Deepfake* diposisikan sebagai variabel bebas (X1) yang diduga berkontribusi terhadap kemampuan deteksi konten *Deepfake*.

Kemampuan Deteksi Konten Digital

Kemampuan deteksi konten digital merupakan kesanggupan seseorang untuk mengidentifikasi dan membedakan konten asli dari konten hasil manipulasi atau rekayasa digital. Kemampuan ini dipengaruhi oleh berbagai faktor, termasuk tingkat literasi digital, pengalaman terhadap teknologi, serta pengetahuan mengenai ciri-ciri teknis konten manipulatif (Altuncu, Franqueira, & Li, 2024). Semakin realistis hasil manipulasi yang dihasilkan oleh teknologi *AI*, semakin sulit pula bagi individu untuk mendeteksinya secara visual tanpa bantuan pengetahuan atau alat bantu teknis (Payne, 2026). Dalam penelitian ini, kemampuan deteksi diposisikan sebagai variabel terikat (Y1) yang diukur secara empiris melalui uji stimulus, yaitu dengan meminta responden mengidentifikasi konten asli dari pasangan konten asli dan *Deepfake* yang disajikan dalam kuesioner.

METODE PENELITIAN

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen terapan (*applied research*) di bidang *machine learning*. Tujuannya adalah untuk mengembangkan model prediksi permintaan konsumen berdasarkan data historis penjualan Toko Nayla Buah di Kecamatan Ratu Agung, Kota Bengkulu. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada analisis data numerik (jumlah penjualan, tanggal, jenis buah, harga, dan lain-lain) untuk menghasilkan pola dan model prediksi yang terukur [6].

Metode *machine learning* digunakan agar peneliti dapat membangun model yang mampu mempelajari pola permintaan dari data historis secara otomatis dan menghasilkan prediksi permintaan di masa mendatang dengan tingkat akurasi yang lebih baik.

Sampel yang digunakan adalah data penjualan harian buah-buahan utama yang paling sering terjual (misalnya: pisang, apel, jeruk, semangka, melon, dan pepaya). Pengambilan sampel dilakukan dengan teknik *purposive sampling*, yaitu memilih data yang relevan dengan kebutuhan model prediksi dan memiliki kelengkapan atribut yang memadai (tanggal, jenis buah, harga, jumlah penjualan, dan variabel terkait) [7].

Tahapan Penelitian

Penelitian ini menggunakan pendekatan KDD (*Knowledge Discovery in Databases*) untuk menganalisis pola dan memprediksi penjualan pada usaha rumah makan dengan menggunakan algoritma *Random Forest*. Pendekatan ini memungkinkan pengolahan data secara sistematis untuk menghasilkan informasi yang bermanfaat dalam pengambilan keputusan strategis. Tahapan KDD yang diterapkan dalam penelitian ini meliputi [8]:

1. Data Selection

Data yang digunakan dipilih berdasarkan relevansi terhadap tujuan analisis. Variabel yang dimasukkan mencakup harga buah, volume penjualan, tren musiman, promosi, serta faktor eksternal yang berpotensi memengaruhi permintaan. Pemilihan data dilakukan untuk memastikan kualitas dan kelengkapan informasi yang akan digunakan dalam model

2. Data Preprocessing

Data mentah yang telah dikumpulkan diproses untuk membersihkan anomali, mengatasi nilai yang hilang (*missing values*), serta menghilangkan duplikasi. Pada tahap ini, normalisasi data dan

penyelarasan format dilakukan agar data siap digunakan untuk analisis lebih lanjut. Kualitas data yang baik merupakan dasar untuk menghasilkan model prediksi yang akurat.

3. Data Transformation

Data yang telah dipra-proses kemudian diubah ke dalam format yang lebih sesuai untuk analisis data mining. Langkah ini mencakup teknik *feature engineering* untuk membuat variabel baru yang relevan, serta teknik *feature selection* menggunakan analisis *Feature Importance* dari algoritma *Random Forest* untuk memilih variabel yang paling signifikan, seperti *Average Sales* dan *Total Sales*

4. Data Mining

Tahap ini merupakan inti dari proses KDD, di mana algoritma *Random Forest* digunakan untuk membangun model prediksi. Model dilatih menggunakan data historis yang telah diproses untuk mengenali pola dan hubungan antara variabel. Parameter model, seperti jumlah pohon (*number of trees*) dan kedalaman maksimum (*max depth*), dioptimalkan untuk menghasilkan prediksi dengan akurasi tinggi

5. Evaluation

Pola yang ditemukan melalui model dievaluasi untuk mengukur performanya. Metrik evaluasi yang digunakan, seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan tingkat akurasi, digunakan untuk menilai efektivitas model dalam memprediksi penjualan.

6. Knowledge Presentation

Hasil penelitian disajikan dalam bentuk laporan yang mencakup temuan utama, visualisasi data (grafik atau tabel), serta interpretasi hasil analisis.

Analisis Masalah

Penelitian ini menggunakan pendekatan *machine learning* dengan algoritma *Random Forest* yang diterapkan menggunakan bahasa pemrograman *Python*, untuk memprediksi permintaan konsumen di toko buah "Nayla". Proses analisis dimulai dengan preprocessing data, termasuk pembersihan, penanganan data hilang, dan normalisasi, untuk memastikan algoritma dapat berjalan optimal. Variabel yang digunakan meliputi jenis buah, jumlah terjual, waktu pembelian, dan metode pembayaran, yang diproses agar tidak ada anomali atau duplikasi. Setelah data siap, algoritma *Random Forest* diterapkan dengan membuat beberapa *decision trees* dari data yang tersedia, yang digabungkan untuk menghasilkan prediksi yang lebih akurat. Secara matematis, algoritma *Random Forest* bekerja dengan menggabungkan berbagai subsamples dari dataset untuk menghasilkan *ensemble* dari *decision trees*, dengan rumus sebagai berikut:

$$f(x) = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

Di mana:

N adalah jumlah *decision trees*

$T_i(x)$ adalah prediksi dari setiap *tree* untuk input x .

Dataset dibagi menjadi dua bagian, yaitu *training set* dan *testing set*. Model dilatih menggunakan *training set* dan diuji pada *testing set* untuk mengevaluasi akurasi prediksi.

Evaluasi Model

Kinerja algoritma *Random Forest* dievaluasi menggunakan beberapa metrik evaluasi, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Metrik ini digunakan untuk mengevaluasi seberapa baik model dalam memprediksi tren penjualan berdasarkan data historis. Rumus yang digunakan untuk metrik-metrik tersebut adalah sebagai berikut [9]:

1. Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

$$Precision = \frac{TP}{TP + FP}$$

3. Recall

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Di mana:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

HASIL DAN PEMBAHASAN

Hasil

Hasil penelitian menunjukkan bahwa faktor jenis buah, faktor hari, promosi, dan cuaca memiliki pengaruh terhadap jumlah permintaan buah oleh konsumen. Variabel-variabel tersebut mampu digunakan sebagai dasar dalam melakukan prediksi permintaan buah pada toko buah. Hasil prediksi dari kedua model dapat digunakan sebagai dasar perencanaan stok dan pengambilan keputusan operasional. Toko dapat mengurangi risiko kerugian akibat kelebihan stok maupun kekurangan stok dengan memperkirakan permintaan harian secara lebih akurat. Secara keseluruhan, Random Forest Regressor adalah model yang paling direkomendasikan untuk diterapkan dalam memprediksi permintaan buah di Toko Nayla Buah karena memiliki tingkat akurasi yang lebih tinggi, stabil, dan mampu memproses variasi data secara lebih efektif.

Pembahasan

Sub bab Data yang digunakan merupakan data simulasi yang mewakili penjualan harian berbagai jenis buah selama 6 bulan dari bulan Januari 2025 sampai dengan Juni 2025. Dataset terdiri dari 1.448 data transaksi yang memuat variabel tanggal transaksi, jenis buah, jumlah terjual, harga, faktor hari, promosi atau diskon, serta kondisi cuaca.

Analisis Regresi Linear

Berdasarkan hasil analisis regresi linear pada data prediksi permintaan konsumen toko buah, diperoleh persamaan regresi linear sebagai berikut:

$$Y = 52,734 - 0,0008X_1 + 1,245X_2 + 4,318X_3 + 6,527X_4 - 2,114X_5$$

Keterangan:

Y = Jumlah permintaan/jumlah buah terjual

X1 = Harga buah

X2 = Jenis buah (hasil encoding)

X3 = Faktor hari

X4 = Promosi/diskon

X5 = Cuaca

52,734 = Konstanta regresi

Penjelasan persamaan regresi:

- Nilai konstanta sebesar 52,734 menunjukkan bahwa apabila seluruh variabel independen dianggap konstan atau bernilai nol, maka jumlah permintaan buah diperkirakan sebesar 52,734 unit.

- b. Koefisien harga sebesar -0,0008 menunjukkan bahwa setiap kenaikan harga sebesar satu satuan akan menyebabkan penurunan jumlah permintaan buah sebesar 0,0008 unit, dengan asumsi variabel lain tetap. Hal ini menunjukkan hubungan negatif antara harga dan permintaan konsumen.
- c. Koefisien jenis buah sebesar 1,245 menunjukkan bahwa perbedaan jenis buah mempengaruhi jumlah permintaan sebesar 1,245 unit.
- d. Koefisien faktor hari sebesar 4,318 menunjukkan bahwa kondisi hari seperti akhir pekan atau hari libur dapat meningkatkan permintaan buah sebesar 4,318 unit dibandingkan hari biasa.
- e. Koefisien promosi sebesar 6,527 menunjukkan bahwa adanya promosi atau diskon mampu meningkatkan jumlah permintaan buah sebesar 6,527 unit.
- f. Koefisien cuaca sebesar -2,114 menunjukkan bahwa kondisi cuaca tertentu, seperti hujan atau mendung, cenderung menurunkan jumlah permintaan buah sebesar 2,114 unit.

Contoh perhitungan prediksi:

Misalnya diketahui:

Harga buah = 25.000

Jenis buah = Apel (kode 0)

Faktor hari = Hari Libur (kode 2)

Promosi = Ya (kode 1)

Cuaca = Cerah (kode 0)

Maka:

$$Y = 52,734 - 0,0008(25000) + 1,245(0) + 4,318(2) + 6,527(1) - 2,114(0)$$

$$Y = 52,734 - 20 + 0 + 8,636 + 6,527 - 0$$

$$Y = 47,897$$

Hasil prediksi menunjukkan bahwa jumlah permintaan buah diperkirakan sebesar 48 unit penjualan. Setelah model Regresi Linear selesai dilakukan, tahap berikutnya adalah melakukan evaluasi model untuk mengetahui tingkat akurasi prediksi. Pada penelitian ini digunakan tiga metode evaluasi, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R-Square (R^2).

Tabel 1. Hasil evaluasi model Regresi Linear

Metode	MAE	RMSE	R^2
Regresi Linear	9,006	11,157	0,235

Nilai MAE sebesar 9,006 menunjukkan bahwa rata-rata kesalahan prediksi model sebesar 9 unit penjualan. Nilai RMSE sebesar 11,157 menunjukkan bahwa model masih memiliki error yang cukup besar. Sedangkan nilai R^2 sebesar 0,235 menunjukkan bahwa model hanya mampu menjelaskan 23,5% variasi permintaan konsumen

Analisis Random Forest

Random Forest merupakan metode ensemble learning yang bekerja dengan membangun banyak decision tree kemudian menggabungkan seluruh hasil prediksi.

Rumus Random Forest:

$$\hat{Y} = (T1(X)+T2(X)+...+Tn(X))/n$$

Keterangan:

$\hat{Y}$ = hasil prediksi akhir

$T_i(X)$ = hasil prediksi tiap pohon

n = jumlah pohon keputusan

Pada penelitian ini digunakan 100 decision tree.

Sebagai contoh:

Tabel 2. Hasil Prediksi

Tree	Hasil Prediksi
Tree 1	45
Tree 2	43
Tree 3	47
Tree 4	44
Tree 5	46

Maka:

$$\hat{Y} = (45+43+47+44+46)/5$$

$$\hat{Y} = 225/5$$

$$\hat{Y} = 45$$

Hasil prediksi akhir Random Forest adalah 45 kilogram.

Setelah dilakukan proses pelatihan (training) dan pengujian (testing) menggunakan metode Random Forest, langkah selanjutnya adalah melakukan evaluasi model untuk mengetahui tingkat akurasi prediksi permintaan buah oleh konsumen. Evaluasi model dilakukan dengan menggunakan tiga indikator pengukuran, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). Metode Random Forest dipilih dalam penelitian ini karena mampu mengolah data dengan pola hubungan yang kompleks dan dapat meningkatkan akurasi prediksi melalui pembentukan banyak pohon keputusan (decision tree). Setiap pohon keputusan akan menghasilkan prediksi, kemudian seluruh hasil prediksi digabungkan untuk memperoleh nilai prediksi akhir yang lebih stabil dan akurat.

Berdasarkan hasil pengujian menggunakan data penelitian, diperoleh hasil evaluasi model Random Forest sebagai berikut:

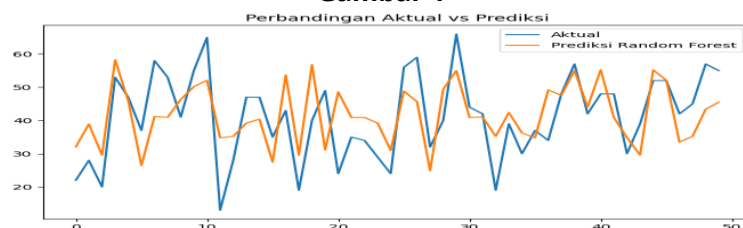
Tabel 3. Hasil Evaluasi Model Random Forest

Metode	MAE	RMSE	R^2
Random Forest	8,293	9,977	0,388

Interpretasi:

Nilai MAE sebesar 8,293 menunjukkan bahwa rata-rata kesalahan prediksi model hanya sekitar 8 unit penjualan. Nilai RMSE sebesar 9,977 menunjukkan tingkat error lebih kecil dibanding regresi linear. Sedangkan nilai R^2 sebesar 0,388 menunjukkan bahwa model mampu menjelaskan 38,8% variasi permintaan konsumen. Metode ini menunjukkan kemampuan prediksi yang sangat baik karena mampu menangkap pola kompleks dari data penjualan buah.

Gambar 1



Grafik hasil prediksi menunjukkan bahwa garis prediksi Random Forest cenderung mengikuti pola data aktual. Hal ini menunjukkan bahwa model mampu melakukan forecasting permintaan buah dengan cukup baik. Pada beberapa titik terdapat selisih antara nilai aktual dan prediksi. Hal tersebut dapat terjadi karena adanya faktor lain di luar penelitian seperti perubahan perilaku konsumen, stok barang, maupun kondisi ekonomi

Analisis KNN

KNN bekerja dengan mencari sejumlah tetangga terdekat berdasarkan jarak Euclidean.

Rumus Euclidean Distance:

$$d(x,y)=\sqrt{((x1-y1)^2+(x2-y2)^2+...)}$$

Keterangan:

$d(x,y)$ = jarak antar data

x_i = data training

y_i = data testing

Pada penelitian ini digunakan nilai $K = 5$.

Contoh Perhitungan KNN, misalnya terdapat dua data:

Tabel 4. Contoh Perhitungan KNN

Variabel	Data 1	Data 2
Harga	25	20
Promosi	1	0
Cuaca	0	1

Maka:

$$d = \sqrt{((25-20)^2+(1-0)^2+(0-1)^2)}$$

$$d = \sqrt{25+1+1}$$

$$d = \sqrt{27}$$

$$d = 5,19$$

Semakin kecil nilai jarak maka semakin dekat hubungan antar data.

Setelah dilakukan proses pelatihan dan pengujian data menggunakan metode K-Nearest Neighbor (KNN), tahap selanjutnya adalah melakukan evaluasi model untuk mengetahui tingkat kemampuan model dalam memprediksi permintaan buah oleh konsumen. Evaluasi dilakukan menggunakan beberapa indikator pengukuran, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2).

Metode KNN merupakan salah satu algoritma machine learning yang bekerja dengan mencari sejumlah tetangga terdekat berdasarkan tingkat kemiripan data. Dalam penelitian ini, metode KNN digunakan untuk memprediksi jumlah permintaan buah berdasarkan variabel harga, jenis buah, faktor hari, promosi, dan kondisi cuaca. Prediksi dilakukan dengan membandingkan data baru terhadap data sebelumnya yang memiliki karakteristik paling mirip. Berdasarkan hasil pengujian model KNN pada data penelitian, diperoleh hasil evaluasi sebagai berikut:

Tabel 5. Hasil Evaluasi

Metode	MAE	RMSE	R^2
KNN	8,421	10,295	0,349

Hasil evaluasi tersebut menunjukkan bahwa metode KNN mampu menghasilkan tingkat prediksi yang cukup baik dalam memprediksi permintaan buah oleh konsumen. Nilai MAE sebesar 8,421 menunjukkan bahwa rata-rata kesalahan prediksi model sebesar 8 unit penjualan. Nilai RMSE sebesar 10,295 menunjukkan tingkat error model dalam memprediksi data penelitian. Sedangkan nilai R^2 sebesar 0,349 menunjukkan bahwa model KNN mampu menjelaskan variasi data permintaan buah sebesar 34,9%.

Berdasarkan hasil tersebut, metode KNN memiliki performa yang lebih baik dibandingkan Regresi Linear, namun masih berada di bawah metode Random Forest. Hal ini menunjukkan bahwa metode KNN cukup efektif digunakan dalam analisis prediksi permintaan konsumen karena mampu mengenali pola kemiripan antar data transaksi penjualan buah.

Perbandingan Hasil Seluruh Model

Untuk mengetahui metode prediksi yang memiliki performa terbaik dalam memprediksi permintaan konsumen pada toko buah, maka dilakukan perbandingan hasil evaluasi dari ketiga metode machine learning yang digunakan dalam penelitian ini, yaitu Regresi Linear, Random Forest, dan K-Nearest Neighbor (KNN).

Perbandingan dilakukan menggunakan beberapa indikator evaluasi model, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). Nilai MAE dan RMSE digunakan untuk mengukur tingkat kesalahan prediksi model, dimana semakin kecil nilai error maka semakin baik kemampuan model dalam melakukan prediksi. Sedangkan nilai R^2 digunakan untuk mengetahui seberapa besar kemampuan model dalam menjelaskan variasi data permintaan konsumen. Semakin besar nilai R^2 , maka semakin baik performa model prediksi. Berdasarkan hasil pengujian dan evaluasi yang telah dilakukan, diperoleh perbandingan performa ketiga metode sebagai berikut:

Tabel 6. Perbandingan Performa Ketiga Metode

Metode	MAE	RMSE	R^2	Keterangan
Regresi Linear	9,006	11,157	0,235	Kurang optimal
Random Forest	8,293	9,977	0,388	Terbaik
KNN	8,421	10,295	0,349	Cukup baik

Berdasarkan hasil perbandingan evaluasi model pada tabel di atas, dapat diketahui bahwa masing-masing metode memiliki tingkat kemampuan prediksi yang berbeda dalam memprediksi permintaan konsumen pada toko buah.

Metode Regresi Linear memperoleh nilai MAE sebesar 9,006 dan RMSE sebesar 11,157 dengan nilai R^2 sebesar 0,235. Hasil tersebut menunjukkan bahwa kemampuan model Regresi Linear dalam memprediksi permintaan buah masih tergolong kurang optimal. Nilai error yang cukup besar menunjukkan bahwa hasil prediksi masih memiliki selisih yang cukup tinggi dibandingkan data aktual. Selain itu, nilai R^2 sebesar 0,235 menunjukkan bahwa model hanya mampu menjelaskan variasi data sebesar 23,5%, sedangkan sisanya dipengaruhi oleh faktor lain di luar model penelitian. Hal ini terjadi karena Regresi Linear hanya mampu menangkap hubungan yang bersifat linear antar variabel.

Metode Random Forest memperoleh hasil terbaik dengan nilai MAE sebesar 8,293 dan RMSE sebesar 9,977 serta nilai R^2 sebesar 0,388. Nilai error yang lebih kecil dibandingkan metode lainnya menunjukkan bahwa Random Forest memiliki tingkat kesalahan prediksi yang lebih rendah. Selain itu, nilai R^2 sebesar 0,388 menunjukkan bahwa model mampu menjelaskan variasi data permintaan buah sebesar 38,8%. Hasil tersebut menunjukkan bahwa Random Forest lebih mampu mengenali pola hubungan yang kompleks antara variabel harga, jenis buah, faktor hari, promosi, dan cuaca terhadap jumlah permintaan konsumen. Oleh karena itu, metode Random Forest dinilai sebagai metode terbaik dalam penelitian ini.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian mengenai *"Prediksi Permintaan Konsumen di Toko Nayla Buah Menggunakan Metode Random Forest Regressor dan K-Nearest Neighbor (KNN)"*, dapat disimpulkan beberapa hal sebagai berikut:

Hasil penelitian menunjukkan bahwa faktor jenis buah, faktor hari, promosi, dan cuaca memiliki pengaruh terhadap jumlah permintaan buah oleh konsumen. Variabel-variabel tersebut mampu digunakan sebagai dasar dalam melakukan prediksi permintaan buah pada toko buah. Proses pengolahan data dilakukan melalui tahapan encoding terhadap data kategorik agar dapat diproses oleh algoritma machine learning. Variabel jenis buah, faktor hari, promosi, dan cuaca berhasil diubah ke dalam bentuk numerik sehingga memudahkan proses analisis dan pemodelan data.

Metode KNN memberikan hasil yang cukup baik, namun tingkat akurasi sedikit lebih rendah. KNN cenderung kurang responsif terhadap perubahan kondisi seperti kenaikan harga atau perbedaan cuaca yang signifikan karena model ini hanya mengandalkan kedekatan jarak antar data. Hasil prediksi dari kedua model dapat digunakan sebagai dasar perencanaan stok dan pengambilan keputusan operasional. Toko dapat mengurangi risiko kerugian akibat kelebihan stok maupun kekurangan stok dengan memperkirakan permintaan harian secara lebih akurat. Secara keseluruhan, Random Forest Regressor adalah model yang paling direkomendasikan untuk diterapkan dalam memprediksi permintaan buah di Toko Nayla Buah karena memiliki tingkat akurasi yang lebih tinggi, stabil, dan mampu memproses variasi data secara lebih efektif.

Saran

Berdasarkan penelitian dan hasil analisis, beberapa saran yang dapat diberikan antara lain:

Pemilik Toko Nayla Buah sebaiknya menggunakan model prediksi Random Forest secara rutin untuk merencanakan jumlah stok harian. Dengan prediksi yang lebih akurat, toko dapat meminimalkan kerusakan buah dan memaksimalkan keuntungan. Pengumpulan data harus dilakukan secara lebih sistematis dan terstruktur, terutama terkait harga, kondisi cuaca, promosi, dan tren pembelian. Data yang lebih lengkap akan meningkatkan kualitas prediksi model machine learning.

Toko disarankan melakukan pencatatan digital menggunakan aplikasi atau spreadsheet sehingga mempermudah proses analisis dan pengolahan model prediksi di masa mendatang. Penelitian selanjutnya dapat menambahkan variabel tambahan, seperti hari libur, event tertentu, promosi toko, atau jenis buah tertentu untuk meningkatkan akurasi model dan memperluas cakupan prediksi.

DAFTAR PUSTAKA

- Aletheia Rabbani, (2019), "Pengertian Prediksi Dan Bentuknya - Sosial79."
- Alvin Darmansyah. (2026). Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Prediksi Kelulusan Mahasiswa Berdasarkan Nilai. *JRIIN : Jurnal Riset Informatika dan Inovasi Volume 3, No. 10*
- Amma, M., Permana, S. E., & Rinaldi, A. R. (2023). Penerapan Dan Pelatihan Perpustakaan Digital Desa Cangkring Kabupaten Cirebon. *AMMA: Jurnal Pengabdian Masyarakat*, 2(8).
- Anugerah, A. G. (2018). Klasifikasi Tingkat Keganasan Kanker Paru-paru pada Citra Computed Tomography (CT) Scan Menggunakan Metode Convolutional Neural Network= Classification Level Of Cancer Rate Rations In Computed Tomography (Ct) Scan Using Convolutional Neural Network Method (Doctoral dissertation, Institut Teknologi Sepuluh Nopember)

- Apriliana Sari . (2025). Prediksi Kebutuhan Stok Barang Menggunakan Algoritma Random Forest Untuk Meningkatkan Efisiensi Penjualan. *Joisie (Journal Of Information Systems And Informatics Engineering)* Vol. 9, No.2, Desember 2025, Hlm 339-351 Received: 24 Juni 2025
- Asohi,Y., (2020), "Implementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan", Universitas Bina Darma
- Breiman, L. (2021). *Random forests*. Machine Learning, 45(1), 5–32.
- Chopra, S., & Meindl, P. (2021). *Supply Chain Management: Strategy, Planning, and Operation* (7th ed.). Pearson
- Defiyanti Sofi, M. K. (2017). Analisis dan Prediksi Kinerja Mahasiswa. Menggunakan Teknik *Data Mining*. Syntak, 2, 1–10
- Farah Nur Farida (2025), Penerapan Model Prediksi Penjualan Pada Usaha Rumah Makan Menggunakan Algoritma *Random Forest*. *Jusim (Jurnal Sistem Informasi Musirawas)*. <https://Doi.Org/10.32767/jusim.V7i2.1701>
- Hahn, T. Von (2020). *Feature Engineering and End-to-End Deep Learning in Tool Wear Monitoring*. Queen's University. (Tesis/Penelitian mengenai deep learning)
- Harun, R., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *Jurnal Manajeme Informatika & Sistem Informasi*, 3(1), 8–15
- Hudaningsih, N., Utami, S. F., & Jabbar, W. A. A. (2020). Perbandingan Peramalan Penjualan Produk Aknil PT. Sunthi Sepuri Menggunakan Metode Single Moving Average dan Single Exponential Smoothing. *Jurnal JINTEKS*, 2(1), 15-22
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- Jet Homepage, S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa, *IJCIT (Indonesian Journal on Computer and Information Technology)*.
- Liparas, D & Dimitriadis, S. I. (2018). How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for AD from ADNI database. *Journal of Neuroscience Methods*, 302, 104-115
- Louppe, G. (2016). *Understanding Random Forests: From Theory to Practice* (Tesis PhD). Université de Liège
- Lumbantoruan R. and P. S. J. Kennedy, (2019). Analisis data mining dan warehousing, *Buletin Ekonomi*, vol. 19, no. 1, pp. 66-76,
- Marisa, F., Maukar, A. L., & Akhirza, T. M. (2021). *Data Mining Konsep Dan Penerapannya*. Deepublish.
- Muhammad Syahrul Efensi. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *Journal Resolusi: Rekayasa Teknik Informatika Dan Informasi* Vol. 5 No. 01
- Setiadi Nugroho J.. (2019). *Perilaku Konsumen: Perspektif Kontemporer Pada Motif, Tujuan, Dan Keinginan Konsumen*. Jakarta. Prenadamedia Group.
- Shalev-Shwartz, S., & Ben-David, S. (2019). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press
- Subekti Rino, dkk. (2024). *Transformasi Digital*. Jambi. Sonpedia Publishing
- Sulianta, F., (2023). *Basic Data Mining From A To Z*. Bandung. Feri Sulianta.
- Syahrudin N and T. Kurniawan, (2018). "Input Dan Output Pada Bahasa Pemrograman *Python*," J. Dasar Pemrograman *Python* Stmik, No. January, Pp. 1–7, 2018.
- Tjiptono, Fandy. 2018. *Manajemen Pemasaran dan Analisa Perilaku Konsumen*. Yogyakarta: BPFE