JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

PERBANDINGAN METODE *K-NEAREST NEIGHBORS (KNN)* DAN *RANDOM FOREST* UNTUK PREDIKSI PENDAPATAN *DRIVER SHOPEEFOOD* DI KUDUS

Fiki Ady Ansyah*[1)], Alif Catur Murti[2)], Ratih Nindyasari[3)]
1. Prodi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus, Indonesia
2. Prodi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus, Indonesia
3. Prodi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus, Indonesia

Article Info
Kata Kunci: *K-Nearest Neighbors (KNN), Random Forest,* Prediksi Pendapatan Driver, ShopeeFood Kudus

Keywords: *K-Nearest Neighbors (KNN), Random Forest, Driver Income Prediction*, *ShopeeFood* Kudus

Article history:
Received 11 January 2025
Revised 16 February 2025
Accepted 25 February 2025
Available 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7335

* Corresponding author.
E-mail address:
fikiadyansyah@gmail.com

ABSTRAK
Perkembangan layanan pesan antar makanan, seperti ShopeeFood, telah menciptakan kebutuhan akan sistem prediksi pendapatan yang akurat untuk mendukung pengambilan keputusan strategis. Penelitian ini membandingkan performa algoritma K-Nearest Neighbors (KNN) dan Random Forest dalam memprediksi pendapatan driver ShopeeFood di Kudus. Dataset primer yang digunakan mencakup 1.443 riwayat perjalanan dengan variabel seperti jumlah order, ongkir, rating, dan ulasan restoran. Hasil menunjukkan bahwa Random Forest lebih unggul dengan nilai MAPE sebesar 5.68%, RMSE sebesar 1,715.39, dan R^2 sebesar 0.74, dibandingkan KNN dengan MAPE sebesar 11.51%, RMSE sebesar 2,290.90, dan R^2 sebesar 0.53. Random Forest terbukti lebih akurat dalam menangkap pola non-linear antar variabel, menjadikannya metode yang lebih andal untuk prediksi pendapatan. Penelitian ini memberikan rekomendasi penggunaan Random Forest sebagai alat utama untuk memaksimalkan efisiensi operasional driver ShopeeFood.

ABSTRACT
The development of food delivery services, such as ShopeeFood, has created the need for an accurate income prediction system to support strategic decision making. This study compares the performance of the K-Nearest Neighbors (KNN) and Random Forest algorithms in predicting the income of ShopeeFood drivers in Kudus. The primary dataset used includes 1,443 trip histories with variables such as number of orders, shipping, ratings, and restaurant reviews. The results show that Random Forest is superior with a MAPE value of 5.68%, RMSE of 1,715.39, and R^2 of 0.74, compared to KNN with a MAPE of 11.51%, RMSE of 2,290.90, and R^2 of 0.53. Random Forest proved to be more accurate in capturing non-linear patterns between variables, making it a more reliable method for income prediction. This research provides a recommendation to use Random Forest as the main tool to maximize the operational efficiency of ShopeeFood drivers.

I. PENDAHULUAN

Transportasi memiliki peranan penting dalam mendukung pembangunan ekonomi, sosial, dan mobilitas masyarakat. Sebagai kebutuhan turunan dari aktivitas ekonomi, ketersediaan jasa transportasi berkorelasi positif dengan perkembangan ekonomi serta alokasi sumber daya yang lebih optimal. Berdasarkan data [1], pasar layanan pesan-antar makanan online di Indonesia mencapai nilai US$ 3.7 miliar pada tahun 2023, dengan pertumbuhan tahunan (CAGR 2023-2027) sebesar 8.44%. Dalam hal pangsa pasar, *GrabFood* memimpin dengan 48% pangsa pasar, diikuti *GoFood* dengan 35%, dan *ShopeeFood* dengan 15% pada kuartal IV 2023. Sejak awal 2021, *ShopeeFood* telah mencatat pertumbuhan signifikan dengan peningkatan jumlah merchant sebesar 127% *year-on-year* pada 2023, serta peningkatan volume transaksi sebesar 85% dibandingkan tahun sebelumnya. Ekspansi layanan ini ditandai dengan penggandaan berbagai industri makanan dan minuman sebagai mitra, serta implementasi strategi pemasaran agresif melalui program diskon dan voucher untuk memperluas jangkauan pasar.[2] *PT Shopee Internasional Indonesia*, yang mulai beroperasi pada April 2020, awalnya fokus pada layanan *e-commerce* penjualan produk dari *seller* ke *customer*, semakin majunya industri teknologi saat ini *Shopee* mulai masuk pada pasar makanan online yang siap bersaing dengan kompetitor serupa seperti *GoFood* dan *Grabfood*,

awalnya *ShopeeFood* hanya memberikan layanan pembelian makanan olahan, minuman ringan, dan kue.[3] Namun, sejak awal 2021, layanan ini memperluas cakupannya dengan menggandeng berbagai industri makanan dan minuman, menarik lebih banyak mitra pengemudi, serta menggelar strategi pemasaran berupa promo diskon dan voucher untuk memperluas jangkauan, khususnya di wilayah Kudus. Upaya ini bertujuan untuk menarik minat konsumen dengan menawarkan harga yang lebih terjangkau dibandingkan ekspektasi mereka.[4] *ShopeeFood* menjadikan sumber daya manusia sebagai faktor utama yang langsung berinteraksi dengan konsumen. Salah satu cara untuk meningkatkan kepercayaan dan loyalitas konsumen adalah dengan memperbaiki kinerja driver *ShopeeFood*. Jika *ShopeeFood* ingin mempertahankan dan meningkatkan kepercayaan serta loyalitas konsumennya, maka perusahaan perlu fokus pada peningkatan kinerja para drivernya. [5] Selain itu, beberapa laporan menunjukkan adanya keluhan dari *driver ShopeeFood* mengenai sistem pembagian order dan insentif yang dianggap kurang adil. Jika masalah ini tidak ditangani dengan baik, hal tersebut dapat memengaruhi kinerja para *driver* secara keseluruhan. Dengan menggunakan metode KNN dan *Random Forest* untuk memprediksi pendapatan, yang dapat membantu manajemen dalam membuat kebijakan berbasis data.[6] metode *K-Nearest Neighbors (KNN)* sederhana, intuitif, dan mampu menangani data baru serta tidak seimbang. Sistem prediksi dikembangkan dengan menggunakan *dataset* historis pendapatan driver *ShopeeFood*. Pengujian menunjukkan bahwa *KNN* berhasil memprediksi pendapatan dan memberikan hasil yang akurat dan relevan.[7] Metode *Random Forest* adalah algoritma *machine learning* berbasis *ensemble* yang efektif dan fleksibel, mampu menangani data kompleks serta mengidentifikasi pola interaksi antar variabel. Dalam penelitian ini, *Random Forest* digunakan untuk membangun sistem prediksi pendapatan driver *ShopeeFood* dengan memanfaatkan dataset historis yang mencakup variabel seperti jumlah order, ongkir, dan faktor lainnya. Keunggulan *Random Forest* terletak pada kemampuannya mengurangi risiko *overfitting* melalui kombinasi prediksi dari beberapa pohon keputusan, sehingga menghasilkan model yang lebih stabil dan akurat. Hasil pengujian menunjukkan bahwa *Random Forest* memberikan prediksi pendapatan yang lebih baik dibandingkan metode lain, dengan akurasi tinggi dan kemampuan menangkap hubungan *non-linear* antara variabel *input* dan *output*. Model ini cocok untuk mendukung pengambilan keputusan strategis baik oleh driver maupun perusahaan.[8] membandingkan kinerja *K-Nearest Neighbors (KNN)* dan *Random Forest* dalam memprediksi pendapatan *driver ShopeeFood* di Kudus. Dataset historis mencakup variabel jumlah order, ongkir, dan data lainnya. *KNN* digunakan untuk menangani data baru, sedangkan *Random Forest* untuk pola data yang kompleks. [9] penelitian sebelumnya yang telah dilakukan untuk mengevaluasi prediksi harga saham menggunakan tiga metode berbeda: regresi linier, regresi Random Forest, dan Multilayer Perceptron (MLP). Hasil penelitian menunjukkan bahwa metode regresi linier menghasilkan nilai Root Mean Square Error (RMSE) terendah, yaitu 0,010, yang mengindikasikan tingkat kesalahan prediksi paling rendah di antara ketiga metode tersebut.[10] Metode *K-Nearest Neighbors (KNN)* dan *Random Forest* digunakan dalam penelitian ini karena memiliki keunggulan unik dalam prediksi pendapatan *driver ShopeeFood*. KNN dipilih karena kesederhanaannya, kemampuan menangani data baru, dan fleksibilitas terhadap data tidak seimbang, sehingga menjadi *baseline* yang baik untuk perbandingan.

Random Forest dipilih karena kemampuannya untuk menangkap hubungan *non-linear* antar variabel, mengurangi risiko *overfitting*, dan menghasilkan model yang stabil pada dataset besar dengan banyak fitur. Hasil pengujian menunjukkan bahwa *Random Forest* memiliki performa yang lebih baik dibandingkan KNN, dengan nilai *Root Mean Square Error (RMSE)* sebesar 1,715.39, lebih rendah dibandingkan KNN yang mencatat *RMSE* sebesar 2,290.90. Selain itu, *Random Forest* juga menghasilkan nilai *Mean Absolute Percentage Error (MAPE)* sebesar 5.68% dan koefisien determinasi (R^2) sebesar 0.74, yang menunjukkan tingkat akurasi prediksi yang lebih tinggi dibandingkan KNN dengan *MAPE* sebesar 11.51% dan R^2 sebesar 0.53. Dengan hasil ini, *Random Forest* terbukti lebih andal untuk mendukung pengambilan keputusan strategis di *ShopeeFood*, khususnya dalam memaksimalkan efisiensi operasional dan pendapatan *driver*.

II. METODE PENELITIAN

2.1 Data Mining

Data mining adalah metode analisis yang bertujuan untuk mengungkap informasi tersembunyi dari kumpulan data berskala besar. Proses ini mencakup berbagai tahapan, seperti pengumpulan, integrasi, ekstraksi, transformasi, hingga analisis data. [11], [12] Salah satu langkah penting dalam data mining adalah konversi data, yang melibatkan penerapan algoritma serta evaluasi untuk mendapatkan hasil yang lebih optimal [11], [12]. Sebagai sebuah proses yang bersifat interaktif dan berulang, data mining bertujuan menciptakan model yang ideal dan mudah dipahami

dari basis data besar. Model ini membantu pengambilan keputusan yang lebih efektif dan dianalisis menggunakan teknik tertentu untuk mendukung pembelajaran yang lebih mendalam [13].

2.2 Regresi

Regresi adalah teknik statistik yang digunakan untuk memodelkan dan menganalisis hubungan antara satu variabel dependen (tak bebas) dengan satu atau lebih variabel independen (bebas). Dalam regresi, tujuan utamanya adalah untuk memprediksi nilai dari variabel dependen berdasarkan nilai-nilai variabel independen. Model regresi yang paling umum adalah regresi linier, yang menggambarkan hubungan linear antara variabel-variabel tersebut [14]. Dalam regresi KNN, prediksi $\hat{y}$ dihitung dengan mengambil rata-rata nilai target y_i dari k tetangga terdekat yang ditemukan dalam ruang fitur berdasarkan metrik jarak tertentu. Rumusnya:

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{1}$$

dengan:

- $\hat{y}$: Prediksi nilai variabel dependen.
- k: Jumlah tetangga terdekat.
- y_i: Nilai target dari tetangga ke-iii.

2.3 *Decision Tree*

Decision Tree adalah model prediktif yang menggunakan struktur pohon untuk membuat keputusan berdasarkan fitur-fitur input. Setiap simpul dalam pohon mewakili sebuah kondisi pada fitur, setiap cabang mewakili hasil dari kondisi tersebut, dan setiap daun mewakili kelas atau nilai output akhir. *Decision Tree* dapat digunakan baik untuk klasifikasi maupun regresi, tergantung pada jenis data yang diolah [15]. Decision Tree menentukan fitur terbaik untuk memisahkan data pada setiap simpul berdasarkan kriteria tertentu, seperti *Information Gain (IG):* Untuk data kategorikal, digunakan dalam pohon klasifikasi. Rumus IG:

$$IG = H(S) - \sum_{i=1}^{n} \frac{|\,S_i\,|}{|\,S\,|} H(S_i)\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{2}$$

Di mana:

- $H(S)$: Entropi dataset S sebelum split.
- S_i: Subset hasil split berdasarkan atribut tertentu.
- $|\,S\,|$: Jumlah total sampel dalam dataset.
- $|\,S_i\,|$: Jumlah sampel dalam subset S_i.

2.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin *non-parametrik* yang digunakan untuk tugas klasifikasi dan regresi. Dalam konteks regresi, KNN memprediksi nilai *output* untuk data baru dengan menghitung rata-rata nilai dari K tetangga terdekat dalam ruang fitur. Proses ini melibatkan pengukuran jarak antara data baru dan data yang ada, biasanya menggunakan metrik seperti jarak *Euclidean*, untuk menentukan tetangga terdekat. KNN regresi sederhana dan efektif, terutama untuk dataset dengan hubungan *non-linear*, namun kinerjanya dapat dipengaruhi oleh pemilihan nilai K dan sensitivitas terhadap skala fitur. [16]

2.5 Random Forest

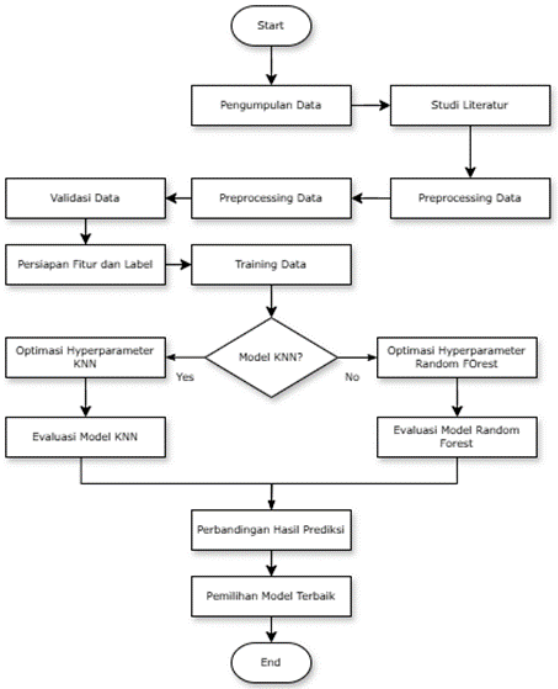
Random Forest, di sisi lain, adalah metode *ensemble* yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi prediksi. Setiap pohon dalam *Random Forest* dibangun dengan menggunakan *subset* acak dari data dan fitur, lalu hasil prediksi dari semua pohon digabungkan dengan cara voting (untuk klasifikasi) atau averaging (untuk regresi) [17]. Pendekatan ini membantu mengurangi *overfitting* dan meningkatkan stabilitas model dibandingkan dengan menggunakan satu pohon keputusan saja [18].

2.6 Tahap Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data, studi literatur, preprocessing data, persiapan fitur dan label, pemisahan data latih dan data uji, pembuatan model KNN dan Random Forest, optimasi

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

hyperparameter, evaluasi model, perbandingan hasil prediksi, hingga pemilihan model terbaik. Alur penelitian ini dapat dilihat pada Gambar 1



Gambar 1 Tahapan Penelitian

2.7 Pengumpulan Data

Penelitian ini menggunakan data primer yang diperoleh dari riwayat perjalanan dan pendapatan oleh 18 *driver ShopeeFood* yang berbeda di Kudus. Data tersebut terdiri dari beberapa atribut yang relevan, seperti Tanggal, Hari, Waktu, Ongkir, Restoran, Double_Order, Orderan, Kategori Restoran, Ulasan Restoran, Rating Restoran, Dan Zona Operasional. Dataset ini mencakup 1.443 data, yang diperoleh selama periode tertentu. Adapun dibawah ini contoh dari data data riwayat perjalanan dan pendapatan driver *ShopeeFood* di Kudus yang digunakan dapat dilihat pada Gambar 2

No	Tanggal	Hari	Waktu	Ongkir	Restoran	Double_Order	Orderan	Kategori	Ulasan	Rating	Zona
1	11/06/2024	Selasa	14:00:00	10400	McDonald	Tidak	1	Masakan Barat	1242	4.1	Kota Kudus
2	11/06/2024	Selasa	15:00:00	7200	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus
3	11/06/2024	Selasa	16:00:00	9600	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus
4	11/06/2024	Selasa	17:00:00	7200	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus
5	11/06/2024	Selasa	18:00:00	7200	Hits Chicken	Tidak	1	Ayam	353	4.3	Kota Kudus
6	11/06/2024	Selasa	19:00:00	16800	Seblak Juara	Tidak	1	Masakan Indonesia	356	4.3	Kota Kudus
7	11/06/2024	Selasa	20:00:00	9600	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus
8	11/06/2024	Selasa	21:00:00	18400	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus
9	11/06/2024	Selasa	22:00:00	8000	McDonald	Tidak	1	Masakan Barat	1242	4.1	Kota Kudus
10	12/06/2024	Rabu	13:00:00	7200	Ayam Geprek Jawi	Tidak	1	Ayam	90	4.1	Bae
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
1435	08/12/2024	Minggu	16:00:00	6400	Kebab Bossgue	Tidak	1	Street Food	39	4.1	Kota Kudus
1436	08/12/2024	Minggu	16:00:00	6400	Nana Crispy Dan Geprek Sadis	Tidak	1	Masakan Indonesia	356	4.3	Kota Kudus
1437	08/12/2024	Minggu	15:00:00	6400	Olive Geprek Expres	Tidak	1	Masakan Indonesia	551	4.7	Bae
1438	08/12/2024	Minggu	15:00:00	24800	Mixue Purwosari Kudus	Tidak	1	Masakan Indonesia	356	4.3	Kota Kudus
1439	08/12/2024	Minggu	14:00:00	6400	Kebab Jawara	Tidak	1	Masakan Indonesia	356	4.3	Kaliwungu
1440	08/12/2024	Minggu	14:00:00	16800	Rm Putra Minang	Ya	2	Masakan Indonesia	120	4.3	Kota Kudus
1441	08/12/2024	Minggu	13:00:00	7200	Ramboo Chicken	Tidak	1	Ayam	2783	4.3	Kota Kudus
1442	08/12/2024	Minggu	12:00:00	6400	Ramboo Chicken	Tidak	1	Ayam	2783	4.3	Kota Kudus
1443	08/12/2024	Minggu	10:00:00	7200	Es Pelopor Kudus	Tidak	1	Masakan Indonesia	356	4.3	Kota Kudus
1444	11/12/2024	Rabu	17:00:00	8000	Mie Gacoan	Tidak	1	Mie	3138	4.2	Kota Kudus

Gambar 2 Data Riwayat Perjalanan dan Pendapatan Driver ShopeeFood di Kudus

2.8 *Preprocessing Data*

Proses preprocessing data merupakan tahapan krusial untuk memastikan kualitas dan kesesuaian data dalam mendukung analisis serta pemodelan yang akurat. Tahapan preprocessing yang diterapkan dalam penelitian ini meliputi beberapa langkah sistematis sebagai berikut:

1. Penghapusan Kolom yang Tidak Relevan : Kolom tertentu, seperti kolom *No*, dihapus karena hanya berfungsi sebagai *identifier* tanpa kontribusi signifikan terhadap analisis data.
2. Transformasi Fitur Kategorikal : Fitur-fitur kategorikal, seperti *Hari*, *Restoran*, dan *Double_Order*, dikonversi menjadi tipe string untuk memastikan kejelasan dalam pemrosesan data. Encoding dilakukan menggunakan teknik *Label Encoding*, yang merupakan metode efisien untuk mengonversi data kategorikal menjadi representasi numerik agar kompatibel dengan algoritma pembelajaran mesin.
3. Ekstraksi Fitur Waktu : Fitur waktu pada kolom *Waktu* diuraikan menjadi dua atribut, yaitu *hours* (jam) dan *time_of_day* (kategori waktu: pagi, siang, malam). Proses ini dilakukan untuk meningkatkan resolusi informasi waktu dalam analisis lebih lanjut.

4. Penanganan *Missing Values* : Nilai kosong (*missing values*) pada data numerik diatasi dengan imputasi menggunakan nilai median, guna meminimalkan pengaruh outlier terhadap distribusi data. Data kategorikal dengan nilai kosong diisi menggunakan modus untuk mempertahankan konsistensi pola kategori.
5. *Feature Engineering* : Dibuat beberapa fitur baru, seperti *Pendapatan_Per_Orderan*, *Order_Rating*, dan *Order_Ulasan*, yang dirancang untuk memberikan wawasan tambahan terhadap perilaku data. Penambahan indikator transaksi akhir pekan (*is_weekend*) dilakukan untuk mengidentifikasi pola transaksi berdasarkan hari operasional tertentu.

Setiap langkah preprocessing ini bertujuan untuk meningkatkan kualitas dan keandalan data, sehingga model analisis yang dihasilkan memiliki performa yang optimal serta interpretasi yang akurat.

Dataset ini kemudian dibagi menjadi dua bagian:

1) Data Latih: 80% dari data untuk membangun dan melatih model.
2) Data Uji: 20% dari data untuk menguji kinerja model.

Tabel 1 berikut menunjukkan atribut dan tipe data yang digunakan untuk penelitian.

TABEL I
VARIABEL DAN TIPE DATA

No	VARIABEL	Tipe Data	Keterangan
1	Hari	*Categorical*	Nama hari transaksi
2	Restoran	*Categorical*	Nama restoran yang melayani pesanan
3	*Double_Order*	*Categorical*	Indikator apakah pesanan adalah double order
4	Orderan	*Numerical*	Jumlah pesanan
5	Rating	*Numerical*	Rating dari restoran
6	Ulasan	*Numerical*	Jumlah ulasan yang diterima restoran
7	Waktu	*Categorical*	Waktu pemesanan dalam format HH:MM
8	Pendapatan	*Numerical*	Total pendapatan dari transaksi
9	Ongkir	*Numerical*	Biaya pengiriman
10	*Hours*	*Numerical*	Jam dari waktu pemesanan
11	*time_of_day*	*Numerical*	Kategori waktu (pagi, siang, malam, dll.)
12	is_weekend	*Binary (0,1)*	Indikator apakah hari termasuk akhir pekan
13	Pendapatan_per_Orderan	*Numerical*	Biaya ongkir per pesanan

Dataset yang telah diproses ini digunakan untuk membandingkan performa dua algoritma, yaitu *K-Nearest Neighbors (KNN)* dan *Random Forest*, dalam memprediksi pendapatan driver *ShopeeFood* di Kudus. Kedua model diuji dan dievaluasi menggunakan metrik seperti *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*.

III. HASIL DAN PEMBAHASAN

3.1 Pemodelan Regresi *K-Nearest Neighbors*

Rumus KNN untuk Prediksi:

$$f(x) = \frac{\sum_{i=1}^{k} w_i \cdot y_i}{\sum_{i=1}^{k} w_i} \tag{3}$$

- x: Data baru yang ingin diprediksi.
- y_i: Nilai target dari tetangga terdekat.
- w_i: Bobot berdasarkan jarak, dihitung dengan rumus:

$$w_i = \frac{1}{\text{distance}(x, x_i)} \tag{4}$$

- k: Jumlah tetangga terdekat.

Pada pemodelan ini:

- Metode validasi: K-Fold Cross-Validation dengan $k = 5$
- Hyperparameter yang diuji: k (jumlah tetangga), metrik jarak (Euclidean, Manhattan, Minkowski).

3.2 Pemodelan *Decision Tree Random Forest*

Rumus Prediksi dengan Random Forest: *Random Forest* adalah metode ensemble yang menggabungkan prediksi dari beberapa decision tree. Rumus prediksi adalah rata-rata dari semua pohon:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (5)$$

- T: Jumlah pohon dalam Random Forest.
- $h_t(x)$: Prediksi dari pohon ke- t.

Proses:

1. Dataset dipecah menjadi banyak subset menggunakan metode bootstrap sampling.
2. Untuk setiap pohon, hanya subset fitur yang digunakan untuk split node.
3. Hasil akhir adalah rata-rata prediksi dari semua pohon (untuk regresi).

Evaluasi Kinerja RF:

- Menggunakan metrik yang sama seperti KNN: MAPE, RMSE, dan R^2

Hyperparameter yang Dicoba:

- $n_{\text{estimators}}$: Jumlah pohon (50, 100, 200).
- max_depth: Kedalaman maksimum pohon (10, 20, None).
- min_samples_split: Minimum sampel untuk split (2, 5, 10).
- max_features: Jumlah fitur maksimum untuk setiap split (sqrt, log2).

3.3 Pengujian

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (6)$$

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (8)$$

3.4 *Tuning Hyperparameter KNN*

Tuning hyperparameter dilakukan pada model *K-Nearest Neighbors (KNN)* untuk menentukan konfigurasi terbaik. *Hyperparameter* yang diuji meliputi jumlah tetangga terdekat (*n_neighbors*: 1 hingga 30), bobot perhitungan jarak (*weights: uniform* dan *distance*), serta metrik jarak yang digunakan (*metric*: *euclidean, manhattan,* dan *minkowski*). Proses tuning dilakukan menggunakan *GridSearchCV* dengan skema *5-Fold Cross-Validation* dan metrik evaluasi *MAPE*. Hasil tuning menunjukkan konfigurasi terbaik KNN adalah *n_neighbors: 5*, *weights: distance*, dan *metric: euclidean*, yang menghasilkan prediksi lebih akurat dengan memanfaatkan pola data secara optimal.

3.5 *Tuning Hyperparameter Random Forest*

Untuk *Random Forest*, hyperparameter yang diuji meliputi jumlah pohon (n_estimators: 50 dan 100), kedalaman maksimum pohon (*max_depth*: 10, 20, dan *None*), jumlah minimum sampel untuk memisahkan node (*min_samples_split*: 5 dan 10), jumlah minimum sampel per daun (*min_samples_leaf*: 2 dan 4), serta jumlah maksimum fitur untuk pemisahan (*max_features*: *sqrt* dan *log2*). Hasil tuning melalui *GridSearchCV* dengan skema 5-*Fold Cross-*

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

Validation menunjukkan konfigurasi terbaik adalah *n_estimators*: 100, *max_depth: None*, *min_samples_split*: 5, *min_samples_leaf*: 2, dan *max_features*: *sqrt*. Konfigurasi ini memungkinkan *Random Forest* menangkap pola data yang kompleks dengan akurasi tinggi dan stabilitas prediksi.

3.6 Hasil Pengujian

TABEL II
PENGUJIAN

Model	TARGET	MAPE	RMSE	R² Score
KNN	Ongkir	11.51%	2,290.90	0.53
Random Forest	Ongkir	5.68%	1,715.39	0.74

Pada Tabel II menyajikan hasil evaluasi performa dua model, yaitu K-Nearest Neighbors (KNN) dan Random Forest, dalam memprediksi biaya ongkir (TARGET). Evaluasi dilakukan menggunakan tiga metrik utama: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2 Score). Berikut penjelasan detail dari hasil yang ditampilkan:

1. *Random Forest* unggul dalam prediksi Ongkir, dengan MAPE yang jauh lebih rendah (5.68%) dibandingkan KNN (11.51%). Ini menunjukkan prediksi yang lebih akurat.
2. KNN cukup baik dalam memprediksi Pendapatan, dengan R^2yang lebih tinggi (0.73) dibandingkan prediksi Ongkir menggunakan KNN (0.53).
3. *Random Forest* lebih stabil dan akurat secara keseluruhan, terutama untuk prediksi Ongkir

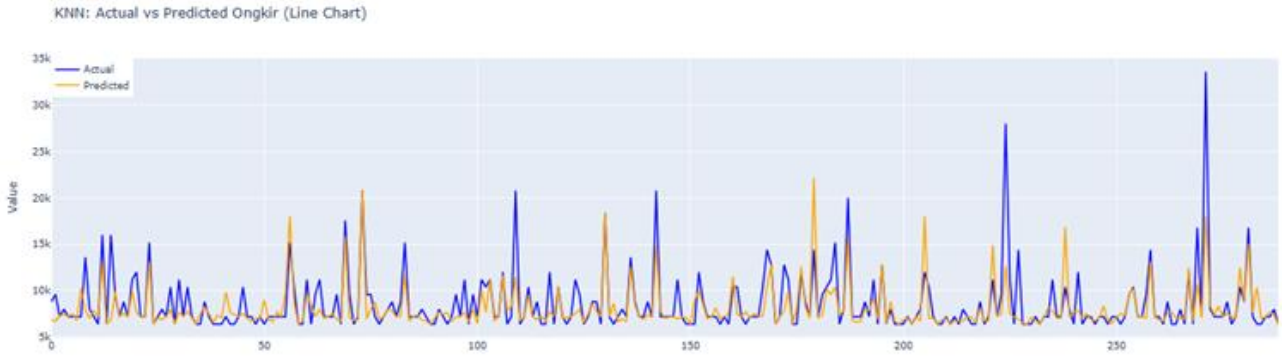
Temuan ini sejalan dengan penelitian yang telah dilakukan [19], yang membandingkan metode regresi linier dan Random Forest dalam memprediksi jumlah transaksi ojek online. Hasil penelitian mereka menunjukkan bahwa regresi linier memiliki kinerja lebih baik dengan nilai RMSE sebesar 31.174, sedangkan nilai RMSE Random Forest mencapai 32.850. Perbedaan hasil ini mungkin disebabkan oleh karakteristik data dan variabel yang digunakan dalam masing-masing penelitian.

Selain itu, penelitian oleh [20] mengklasifikasikan pendapatan driver ShopeeFood menggunakan metode K-Means. Hasil penelitian mereka menunjukkan bahwa pendapatan tertinggi terjadi pada bulan pertama dengan Rp39.000,00, sedangkan pendapatan terendah terjadi pada bulan keempat dengan Rp6.400,00. Meskipun metode yang digunakan berbeda, penelitian ini memberikan wawasan tambahan mengenai variasi pendapatan driver ShopeeFood.

Penelitian dari [21] menemukan hubungan positif antara kenaikan tarif ojek online dan pendapatan pengemudi di Stasiun Bekasi, dengan persamaan regresi $Y' = 37.220 + 0.236X$. Kenaikan tarif menyumbang 3.8% terhadap pendapatan, sementara 96.2% lainnya dipengaruhi faktor lain seperti jumlah pesanan dan jam kerja. Uji t-hitung sebesar 6.296 lebih besar dari t-tabel 2.002, menunjukkan pengaruh yang signifikan. Meskipun pendapatan per perjalanan meningkat, penurunan jumlah pelanggan membatasi dampaknya. Data yang digunakan terbukti reliabel dengan Cronbach Alpha > 0.60. Penelitian ini menyimpulkan bahwa kenaikan tarif berdampak positif namun terbatas, memberikan wawasan untuk kebijakan yang mendukung kesejahteraan pengemudi.

3.7 Perbandingan

Grafik yang ditampilkan pada Gambar 3 dan Gambar 4 adalah perbandingan antara data aktual dan prediksi untuk biaya ongkir (Ongkir) dengan menggunakan dua model: *K-Nearest Neighbors (KNN)* dan *Random Forest* (RF). Berikut adalah analisis hasil grafik:

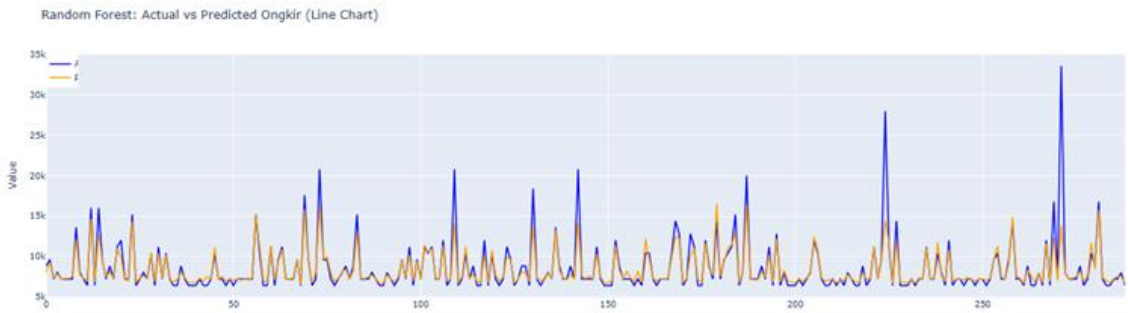


Gambar 3 Perbandingan Data Aktual dan Data Prediksi

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

A. Analisis Grafik: KNN

- Polarisasi Prediksi: Grafik KNN menunjukkan bahwa prediksi (garis oranye) cenderung mengikuti pola data aktual (garis biru). Namun, terdapat beberapa deviasi yang lebih besar di titik-titik tertentu, terutama pada lonjakan nilai Ongkir yang tinggi.
- Kesesuaian: Prediksi KNN cukup akurat untuk nilai Ongkir rendah hingga menengah, tetapi mulai kehilangan akurasi pada nilai Ongkir yang lebih tinggi, seperti pada indeks 100–150 dan 250–300.



Gambar 4 Perbandingan Data Aktual dan Data Prediksi

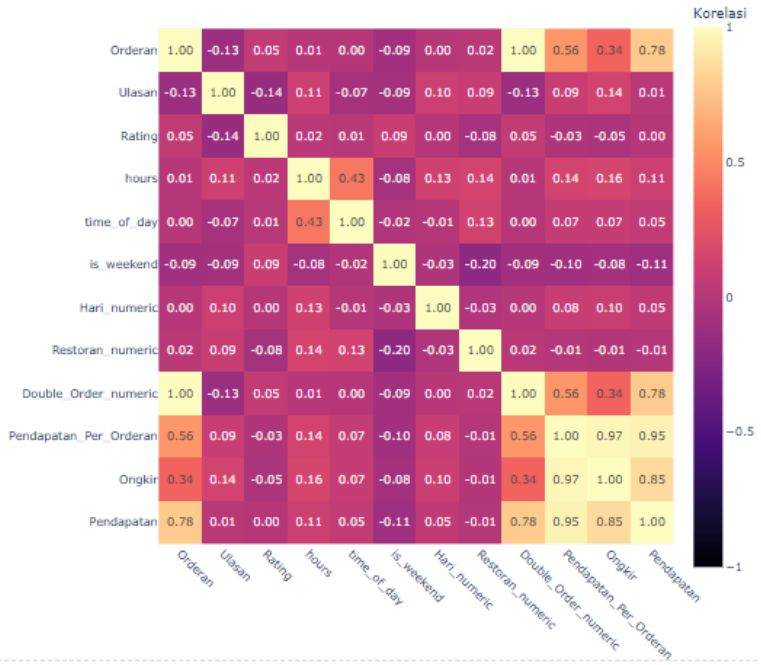
B. Analisis Grafik: *Random Forest*

- Kesesuaian Pola: Grafik *Random Forest* menunjukkan prediksi yang lebih halus dan mengikuti pola data aktual dengan lebih baik dibandingkan KNN, terutama pada titik-titik dengan lonjakan nilai Ongkir yang tinggi.
- Stabilitas Model: Prediksi *Random Forest* menunjukkan stabilitas yang lebih tinggi, dengan lebih sedikit deviasi dari data aktual, baik untuk nilai rendah maupun tinggi.

Tabel III menampilkan perbandingan antara model *K-Nearest Neighbors (KNN)* dan *Random Forest* dalam aspek kesesuaian pola prediksi, stabilitas, dan akurasi berdasarkan hasil prediksi biaya ongkir (Ongkir). Berikut adalah penjelasan rinci setiap aspek:

TABEL III
HASIL PERBANDINGAN DATA AKTUAL DAN DATA PREDIKSI

Aspek	KNN	Random Forest
Kesesuaian Pola	Pola prediksi mengikuti data aktual, tetapi terdapat lebih banyak deviasi.	Pola prediksi sangat mendekati data aktual, bahkan pada lonjakan.
Stabilitas	Kurang stabil, terutama pada nilai Ongkir tinggi.	Stabil dan akurat untuk semua rentang nilai.
Akurasi	Akurasi baik pada nilai rendah, tetapi menurun pada nilai tinggi.	Akurasi tinggi pada semua rentang nilai.



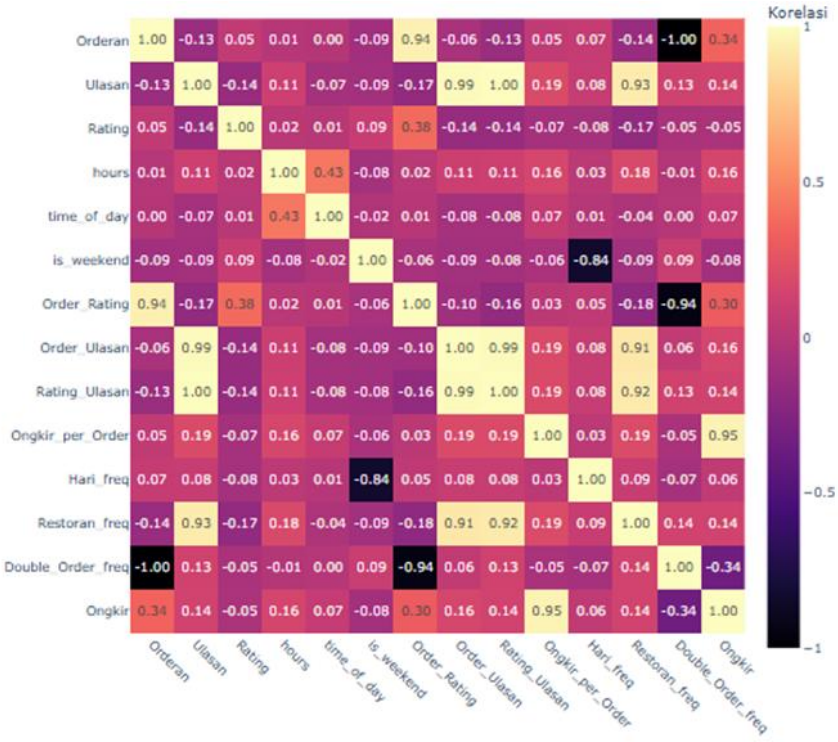
Gambar 5 Heatmap Plot KNN

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

Confuson Matrix pada Gambar 5 Fitur utama yang memiliki hubungan signifikan dengan pendapatan adalah Orderan (0.78), *Double_Order_numeric* (0.78), Pendapatan_Per_Orderan (0.97), dan Ongkir (0.85). Hal ini menunjukkan bahwa jumlah pesanan, pesanan ganda, pendapatan per order, dan ongkir merupakan faktor kunci dalam memengaruhi pendapatan driver *ShopeeFood*. Fitur ini menunjukkan korelasi tinggi, yang mengindikasikan pentingnya memasukkan variabel-variabel tersebut dalam model prediksi untuk meningkatkan akurasi.

Sebaliknya, fitur seperti Rating (-0.05), Ulasan (0.09), dan *is_weekend* (-0.11) memiliki korelasi yang sangat lemah terhadap pendapatan, sehingga kontribusinya terhadap prediksi cukup minim. Fitur waktu, seperti hours (0.43), memiliki pengaruh moderat, yang menunjukkan adanya variasi pendapatan berdasarkan jam operasional tertentu. Namun, *time_of_day* (0.07) memiliki dampak yang hampir tidak signifikan terhadap pendapatan, sehingga fitur ini kurang relevan untuk digunakan dalam model.

Hasil analisis ini memperkuat pentingnya seleksi fitur dalam membangun model prediksi yang akurat. Fitur dengan korelasi tinggi, seperti Orderan dan Pendapatan_Per_Orderan, harus menjadi fokus utama dalam proses pelatihan model, sementara fitur dengan korelasi rendah dapat dipertimbangkan sebagai pelengkap untuk meningkatkan efisiensi model. Selain itu, penggunaan algoritma seperti Random Forest memungkinkan pemanfaatan hubungan non-linear antar variabel secara lebih optimal, dibandingkan dengan algoritma yang hanya bergantung pada korelasi linier seperti KNN.



Gambar 6 Heatmap Plot Random Forest

Pada model *Random Forest*, pola korelasi yang ditemukan lebih rinci karena algoritma ini mampu menangkap hubungan *non-linear* antar variabel. Fitur seperti Order_Rating (0.94), Order_Ulasan (0.99), dan Ongkir_per_Order (0.95) menunjukkan korelasi tinggi terhadap pendapatan, yang mengindikasikan bahwa model *Random Forest* lebih mampu memanfaatkan fitur hasil rekayasa *(feature engineering)*. Hal ini menunjukkan pentingnya fitur-fitur tersebut dalam menghasilkan prediksi pendapatan yang lebih akurat dan relevan.

Sebaliknya, fitur seperti *is_weekend* (-0.08) dan time_of_day (0.07) tetap memiliki korelasi lemah terhadap pendapatan, menunjukkan konsistensi kelemahan fitur ini pada kedua model (KNN dan Random Forest). Korelasi rendah ini mengindikasikan bahwa fitur tersebut memberikan kontribusi minimal terhadap hasil prediksi, sehingga dapat dipertimbangkan untuk tidak menjadi fokus utama dalam pengembangan model.

Hasil analisis pada Gambar 6 juga menunjukkan bahwa Random Forest mampu memanfaatkan hubungan kompleks antar fitur, seperti yang terlihat pada korelasi Restoran_freq (0.14) dan Hari_freq (0.07), yang meskipun tidak signifikan, tetap diolah secara efektif oleh model. Dibandingkan dengan KNN, Random Forest menunjukkan keunggulan dalam memanfaatkan fitur kompleks, yang memberikan hasil prediksi lebih stabil dan akurat, terutama pada fitur yang memiliki korelasi tinggi seperti Order_Rating dan Ongkir_per_Order.

Secara keseluruhan, Gambar 6 memperkuat kesimpulan bahwa Random Forest adalah metode yang lebih efektif dalam memanfaatkan dataset yang kompleks dengan memaksimalkan potensi variabel hasil rekayasa. Fitur dengan

korelasi tinggi perlu menjadi fokus utama, sementara fitur dengan korelasi rendah dapat dipertimbangkan sebagai pelengkap untuk meningkatkan efisiensi model.

Random Forest menunjukkan keunggulan signifikan dalam memanfaatkan hubungan non-linear antar fitur, terutama melalui fitur hasil rekayasa seperti Order_Rating (0.94), Order_Ulasan (0.99), dan Ongkir_per_Order (0.95), yang memiliki korelasi tinggi terhadap pendapatan. Hal ini menunjukkan bahwa model Random Forest mampu mengoptimalkan potensi fitur kompleks untuk menghasilkan prediksi yang lebih akurat. Sebaliknya, KNN hanya mengandalkan hubungan linier langsung antar fitur utama, seperti Orderan (0.78) dan Pendapatan_Per_Orderan (0.97), sehingga kurang mampu menangkap pola yang lebih kompleks dalam dataset.

Fitur seperti *is_weekend* (-0.08 hingga -0.11) dan *time_of_day* (0.07) memiliki korelasi rendah pada kedua model, menunjukkan kontribusi yang minim terhadap prediksi pendapatan. Oleh karena itu, fitur ini dapat dianggap kurang signifikan dan berfungsi lebih sebagai pelengkap.

Dalam pengembangan model prediksi pendapatan *driver ShopeeFood*, beberapa keterbatasan perlu dipertimbangkan untuk interpretasi yang tepat. Model yang dikembangkan mengasumsikan pola pendapatan yang relatif stabil, namun dalam realitasnya, berbagai faktor eksternal dapat mempengaruhi pola tersebut. Keterbatasan utama meliputi:

Pertama, asumsi stasioneritas data yang mengasumsikan stabilitas pola pendapatan sepanjang waktu mungkin tidak sepenuhnya mencerminkan realitas pasar yang dinamis. Perubahan kebijakan perusahaan, fluktuasi permintaan musiman, dan evolusi preferensi konsumen dapat mempengaruhi akurasi prediksi model.

Kedua, dataset yang digunakan memiliki keterbatasan cakupan, hanya meliputi 1.443 transaksi dari 18 driver di wilayah Kudus. Periode pengambilan data yang terbatas juga dapat mempengaruhi kemampuan model dalam menangkap variasi jangka panjang. Beberapa variabel potensial seperti kondisi cuaca dan kepadatan lalu lintas tidak tersedia dalam dataset.

Ketiga, generalisasi model ke wilayah lain perlu dilakukan dengan hati-hati mengingat karakteristik unik setiap daerah. Perbedaan pola konsumsi, infrastruktur transportasi, dan dinamika pasar lokal dapat mempengaruhi aplikabilitas model di luar wilayah Kudus.

Hasil penelitian ini memiliki beberapa implikasi praktis yang dapat diimplementasikan untuk meningkatkan efisiensi operasional *ShopeeFood*. Dalam konteks optimasi operasional, hasil prediksi dapat digunakan untuk mengembangkan sistem penjadwalan driver yang lebih efisien. Sistem ini dapat mengalokasikan sumber daya berdasarkan pola permintaan yang diprediksi, sehingga meningkatkan tingkat penerimaan order dan efisiensi pengiriman. Dari perspektif personalisasi layanan, model prediksi dapat mendukung pengembangan sistem rekomendasi rute yang optimal bagi *driver*. Sistem ini dapat disesuaikan dengan pola permintaan historis dan kondisi *real-time*, memungkinkan driver untuk memaksimalkan pendapatan mereka sambil meminimalkan waktu dan biaya operasional. Skema insentif juga dapat disesuaikan berdasarkan pola pendapatan yang teridentifikasi, menciptakan sistem reward yang lebih adil dan transparan. Dalam hal pengembangan produk, hasil penelitian ini dapat menginformasikan pengembangan fitur-fitur baru dalam aplikasi *ShopeeFood*. Misalnya, implementasi *dashboard* prediksi pendapatan *real-time* untuk driver, sistem notifikasi cerdas untuk peluang pendapatan tinggi, dan *tools* analitik untuk membantu *driver* mengoptimalkan waktu kerja mereka. Pengembangan ini diharapkan dapat meningkatkan kepuasan *driver* dan mendorong retensi jangka panjang.

Pentingnya seleksi fitur utama dengan korelasi tinggi dalam membangun model prediksi yang andal. Algoritma seperti Random Forest unggul karena kemampuannya menangkap hubungan *non-linear* antar variabel, sedangkan KNN lebih cocok digunakan sebagai baseline sederhana untuk analisis awal. Random Forest direkomendasikan untuk memanfaatkan seluruh variabel dalam dataset secara optimal guna menghasilkan prediksi yang lebih akurat dan stabil.

3.8 Prediksi Pendapatan

```
Masukkan data berikut untuk prediksi Ongkir dan Pendapatan:
Masukkan waktu (format HH:MM, misalnya 14:00): 13:00
Masukkan nama hari (misalnya Senin): Rabu
Masukkan nama restoran (misalnya McDonald): Mie Gacoan
Apakah Double Order? (Ya/Tidak): Ya
Masukkan jumlah orderan (integer, misalnya 10): 10
Masukkan rating restoran (float, misalnya 4.1): 4.5
Masukkan jumlah ulasan (integer, misalnya 50): 1000

Hasil Prediksi:
Prediksi Rata-Rata Ongkir: Rp 12,167.74
Total Pendapatan: Rp 121,677.36
```

Gambar 7 Prediksi Pendapatan KNN

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.35-46

```
Masukkan data berikut untuk prediksi Ongkir dan Pendapatan:
Masukkan waktu (format HH:MM, misalnya 14:00): 13:00
Masukkan nama hari (misalnya Senin): Rabu
Masukkan nama restoran (misalnya McDonald): Mie Gacoan
Apakah Double Order? (Ya/Tidak): Ya
Masukkan jumlah orderan (integer, misalnya 10): 10
Masukkan rating restoran (float, misalnya 4.1): 4.5
Masukkan jumlah ulasan (integer, misalnya 50): 1000

Hasil Prediksi:
Prediksi Rata-Rata Ongkir: Rp 12,180.00
Total Pendapatan: Rp 121,799.95
```

Gambar 8 Prediksi Pendapatan Random Forest

Hasil prediksi pendapatan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors (KNN)* menunjukkan perbedaan performa yang signifikan dalam memodelkan data pada skenario yang sama. Pada skenario input dengan waktu transaksi pukul 13:00 pada hari Rabu di restoran Mie Gacoan dengan 10 order, status double order, rating restoran 4.5, dan 1000 ulasan, algoritma *Random Forest* bisa dilihat pada Gambar 7 menghasilkan prediksi rata-rata ongkir sebesar Rp 12,180.00 dan total pendapatan Rp 121,799.95. Sebaliknya, pada Gambar 8 KNN menghasilkan rata-rata ongkir sebesar Rp 12,167.74 dan total pendapatan Rp 121,677.36, dengan selisih masing-masing sebesar Rp 12.26 dan Rp 122.59. Berdasarkan hasil ini, *Random Forest* terbukti lebih unggul dalam memodelkan hubungan antar variabel, sebagaimana didukung oleh nilai MAPE yang lebih rendah sebesar 5.68% dibandingkan KNN sebesar 11.51%, serta nilai RMSE yang lebih baik (1,715.39 dibandingkan 2,290.90). Hal ini menunjukkan kemampuan *Random Forest* untuk menangkap pola non-linear dengan lebih akurat, sementara KNN cenderung lebih cocok untuk data pada rentang rendah hingga menengah namun mengalami deviasi pada nilai yang lebih tinggi.

Keunggulan *Random Forest* terletak pada pendekatan ensemble yang mampu mengurangi risiko overfitting dengan menggabungkan prediksi dari berbagai pohon keputusan, sehingga menghasilkan model yang lebih stabil dan akurat. Prediksi ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa *Random Forest* lebih efektif untuk data yang melibatkan variabel kompleks. Oleh karena itu, *Random Forest* direkomendasikan sebagai metode utama dalam skenario prediksi pendapatan yang memerlukan analisis yang lebih mendalam dan akurasi tinggi. Sementara itu, KNN tetap dapat digunakan sebagai baseline untuk membandingkan performa algoritma lain dalam berbagai skenario analisis.

K-Nearest Neighbors (KNN) sensitif terhadap skala data dan rentan terhadap *outlier*, sehingga membutuhkan normalisasi yang tepat. Model ini kesulitan menangkap pola *non-linear* pada *dataset* kompleks, seperti yang terlihat dari nilai MAPE dan RMSE yang lebih tinggi dibandingkan *Random Forest*. Sebaliknya, *Random Forest* unggul dalam akurasi dan stabilitas, tetapi membutuhkan sumber daya komputasi lebih besar dan waktu pemrosesan yang lebih lama. Selain itu, interpretasi kontribusi fitur pada *Random Forest* kurang intuitif dibandingkan KNN, sehingga dapat menjadi tantangan pada dataset yang sangat besar atau kebutuhan prediksi *real-time*.

IV. KESIMPULAN

Penelitian ini membandingkan performa metode *K-Nearest Neighbors (KNN)* dan *Random Forest* dalam memprediksi pendapatan driver *ShopeeFood* di Kudus. Berdasarkan hasil pengujian menggunakan *dataset* historis yang mencakup 1.443 data perjalanan, *Random Forest* menunjukkan kinerja yang lebih unggul dibandingkan KNN. *Random Forest* menghasilkan nilai *Mean Absolute Percentage Error (MAPE)* sebesar 5.68%, *Root Mean Squared Error (RMSE)* sebesar 1,715.39, dan koefisien determinasi (R^2) sebesar 0.74. Sebaliknya, KNN mencatat MAPE sebesar 11.51%, RMSE sebesar 2,290.90, dan R^2 sebesar 0.53. Keunggulan *Random Forest* terletak pada kemampuannya untuk menangkap hubungan *non-linear* antar variabel, mengurangi risiko *overfitting*, dan menghasilkan model yang lebih stabil, sehingga mampu memberikan prediksi yang lebih akurat.

Hasil penelitian ini menunjukkan bahwa *Random Forest* adalah metode yang lebih andal untuk mendukung pengambilan keputusan strategis di *ShopeeFood*, khususnya dalam upaya memaksimalkan pendapatan dan efisiensi operasional *driver*. KNN, meskipun memiliki performa yang cukup baik dalam memodelkan beberapa variabel, lebih cocok digunakan sebagai baseline untuk membandingkan performa algoritma lainnya. Penelitian di masa depan diharapkan dapat memperluas cakupan dataset dan menyertakan variabel tambahan untuk meningkatkan akurasi model serta memberikan wawasan yang lebih komprehensif.

DAFTAR PUSTAKA

[1] Statista, “Online Food Delivery - Worldwide.”
[2] M. S. Syahbandi, E. S. Tasri, and others, “Pengaruh Kehadiran Transportasi Online Terhadap Pendapatan Transportasi Konvensional Dikota Padang,” *Abstract Of Undergraduate Research, Faculty Of Economics, Bung Hatta University*, vol. 17, no. 3, 2020.

[3] M. Oktaviana, A. D. Nurhalim, and E. Hernawati, "An Analysis Of Go-Food, Grabfood, And Shopeefood Utilization To Improve Customer Loyalty On Home-Based Business Owners In Tangerang City," *Primanomics : Jurnal Ekonomi & Bisnis*, vol. 19, no. 3, pp. 1–10, Sep. 2021, doi: 10.31253/pe.v19i3.631.

[4] I. Vania and R. Simbolon, "Pengaruh Promo ShopeeFood Terhadap Minat Beli Pengguna Shopee (Di Daerah Tangerang Selatan)," *Jurnal Ekonomis*, vol. 14, no. 2b, 2021.

[5] B. Herman, I. Nursyamsi, and Y. Amar, "The Effect Of Financial Compensation, Non Financial Compensation And Work Flexibility On Online Motorcycle Taxi Drivers Performance With Work Satisfaction As Intervening Variables On Gojek In Makassar," *International Journal Of Advanced Research (IJAR)*, vol. 8, no. 7, pp. 632–639, 2020.

[6] M. J. Mahendra, S. Asiyah, and N. Wahyuningtyas, "Pengaruh Work Life Balance, Motivasi Dan Disiplin Kerja Terhadap Kinerja Driver Shopeefood Di Kota Malang (Studi Kasus Pada Driver Shopeefood Di Kota Malang)," *E-JRM: Elektronik Jurnal Riset Manajemen*, vol. 13, no. 02, pp. 152–159, 2024.

[7] E. Ermin, R. Soekarta, L. Jupriadi Fakhri, M. Jundullah, and D. Retno Irianto, "ANALYSIS OF ONLINE TRANSPORTATION CUSTOMER SATISFACTION WITH THE APPLICATION OF TEXT CLASSIFICATION USING KNN METHOD," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 18, no. 2, pp. 244–252, Nov. 2024, doi: 10.35457/antivirus.v18i2.3994.

[8] S. Saadah and H. Salsabila, "Prediksi Harga Bitcoin Menggunakan Metode Random Forest," *Jurnal Komputer Terapan*, vol. 7, no. 1, pp. 24–32, Jun. 2021, doi: 10.35143/jkt.v7i1.4618.

[9] M. A. Pratama, M. Munawaroh, and W. J. Pranoto, "Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda," *TEKTONIK: Jurnal Ilmu Teknik*, vol. 1, no. 2, pp. 172–182, 2024.

[10] E. Fitri and D. Riana, "ANALISA PERBANDINGAN MODEL PREDICTION DALAM PREDIKSI HARGA SAHAM MENGGUNAKAN METODE LINEAR REGRESSION, RANDOM FOREST REGRESSION DAN MULTILAYER PERCEPTRON," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 6, no. 1, pp. 69–78, Apr. 2022, doi: 10.46880/jmika.Vol6No1.pp69-78.

[11] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknologi Informasi*, vol. 8, no. 2, pp. 122–131, Dec. 2022, doi: 10.33372/stn.v8i2.885.

[12] A. Andika, S. Syarli, and C. R. Sari, "DATA MINING KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES," *Journal Peqguruang: Conference Series*, vol. 4, no. 1, p. 423, May 2022, doi: 10.35329/jp.v4i1.2358.

[13] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES," *Technologia : Jurnal Ilmiah*, vol. 13, no. 3, p. 207, Aug. 2022, doi: 10.31602/tji.v13i3.7312.

[14] W. A. L. Sinaga, S. Sumarno, and I. P. Sari, "The Application of Multiple Linear Regression Method for Population Estimation Gunung Malela District," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 1, pp. 55–64, Mar. 2022, doi: 10.55123/jomlai.v1i1.143.

[15] I. Firman Alamsyah, R. Esra, S. Awalia, D. Andi Nohe, J. Matematika, and F. Matematika dan Ilmu Pengetahuan Alam, "Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya Terbitan II," 2022.

[16] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing," *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, pp. 17–29, May 2023, doi: 10.34312/jjps.v4i1.18948.

[17] M. M. Rofi, F. A. Setiawan, and F. Riana, "PERBANDINGAN METODE K-NN DAN RANDOM FOREST PADA KLASIFIKASI MAHASISWA BERPOTENSI DROPOUT," *INFOTECH journal*, vol. 10, no. 1, pp. 84–89, Mar. 2024, doi: 10.31949/infotech.v10i1.8856.

[18] F. Sholekhah, A. D. Putri, R. Rahmaddeni, and L. Efrizoni, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 507–514, Feb. 2024, doi: 10.57152/malcom.v4i2.1249.

[19] D. Pramesti and W. M. Baihaqi, "Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Regresi Linier dan Random Forest," *Generation Journal*, vol. 7, no. 3, pp. 21–30, 2023.

[20] Ariq Azhar Zhafari and Supatman, "Klasifikasi Pendapatan Driver Shopee Food Menggunakan Metode K-Means," *Da'watuna: Journal of Communication and Islamic Broadcasting*, vol. 4, no. 4, Jun. 2024, doi: 10.47467/dawatuna.v4i4.1735.

[21] M. F. Ar Rizqi, P. R. Irawan, and S. Supriyadi, "Pengaruh Kenaikan Tarif Ojek Online Terhadap Pendapatan Driver Ojek Online Di Stasiun Bekasi," *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 2013–2018, Dec. 2024, doi: 10.33395/jmp.v13i2.14363.