

## PREDIKSI HARGA MOTOR BEKAS DI KOTA KUPANG MENGGUNAKAN METODE *RANDOM FOREST*

Mohamad Iqbal Ulumando

Institut Teknologi Alberth Foenay Kupang, Indonesia

[iqbal77ulumando@gmail.com](mailto:iqbal77ulumando@gmail.com)

### ABSTRACT

Used motorcycle prices in the market often fluctuate and are influenced by various factors such as brand, type, year of manufacture, engine capacity, mileage, vehicle condition, and tax status. In Kupang City, the determination of used motorcycle prices is generally still performed manually based on seller estimates or market conditions, which can lead to significant price differences. Therefore, a method is needed to estimate used motorcycle prices more objectively and accurately. This research seeks to develop a model for predicting used motorcycle prices employing the Random Forest algorithm. The dataset used consists of 200 used motorcycle records collected from used motorcycle sales data in Kupang City. The data undergoes a preprocessing stage before being used for model development. Furthermore, the dataset is partitioned into training data and testing to build evaluate prediction. The evaluation results show a Mean Absolute Error (MAE) of Rp.4,418,477, a Mean Squared Error (MSE) of 26,617,157,710,315, and a Root Mean Squared Error (RMSE) of Rp.5,159,181. The findings suggest that the Random Forest predict motorcycle prices with reasonably acceptable error rate, making it a useful approach for estimating used motorcycle prices based on vehicle attributes.

### Keywords:

*Random Forest; Price Prediction; Used Motorcycles; Kupang City.*

### PENDAHULUAN

Perkembangan teknologi informasi dan analisis data telah membawa pengaruh yang besar di berbagai sektor, terutama dalam proses pengambilan keputusan yang berbasis data (Muhammad Fawaati & Rahmatullah, 2022). Salah satu metode yang saat ini banyak dimanfaatkan adalah *machine learning*, yaitu pendekatan yang memungkinkan sistem untuk mempelajari pola dari data sebelumnya sehingga mampu menghasilkan prediksi dengan tingkat akurasi yang lebih baik. (Putri Alesia Nadeak, Irahayu Barimbing, Jois Nursaida Batubara, Jonson Sinaga, 2025). Penerapan *machine learning* telah digunakan dalam berbagai sektor, seperti keuangan, kesehatan, pemasaran, serta analisis harga pasar (Ade Ryan Pratama et al., 2025).

Sepeda motor merupakan salah satu alat transportasi yang paling banyak digunakan oleh masyarakat di Indonesia, termasuk di Kota Kupang (Theresia Felisitas Dena & Franki Yusuf Bisilisin, 2025). Tingginya kebutuhan masyarakat terhadap kendaraan bermotor menyebabkan aktivitas jual beli motor bekas terus meningkat (Wati, 2021). Showroom motor bekas menjadi salah satu tempat yang banyak dipilih masyarakat untuk membeli maupun menjual kendaraan karena menyediakan berbagai pilihan motor dengan kondisi dan harga yang beragam (Masnita Sari, Hera Khairunnisa, 2025).

Penentuan harga motor bekas sering kali dipengaruhi oleh beberapa faktor, seperti merek kendaraan, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi fisik kendaraan, serta status pajak kendaraan (Wildan Hannan et al., 2025). Namun dalam praktiknya, penentuan harga masih banyak dilakukan secara subjektif berdasarkan pengalaman penjual atau kondisi pasar saat itu. Hal ini dapat menyebabkan ketidaksesuaian harga antara nilai kendaraan yang sebenarnya dengan harga yang ditawarkan kepada konsumen (Evasaria Magdalena Sipayung, Cut Fiarni, 2023).

Sebagai upaya untuk mengatasi permasalahan tersebut, diperlukan suatu metode yang dapat membantu memprediksi harga motor bekas secara lebih objektif berdasarkan data yang tersedia (Yanto Herdianto, 2023). Metode yang dapat dimanfaatkan dalam proses ini salah satunya adalah *Random Forest* untuk memprediksi harga motor bekas, dimana penulis sebelumnya juga menggunakan metode *random forest* untuk klasifikasi resiko *DropOut* mahasiswa ITAF Kupang sebagai sistem peringatan dini (Ulumando, 2026). Algoritma *Random Forest* termasuk metode *ensemble learning* yang memanfaatkan sejumlah *decision tree*, lalu mengkonsolidasikan prediksi dari setiap pohon untuk memperoleh hasil akhir keputusan tersebut guna memperoleh hasil prediksi untuk memperoleh prediksi yang lebih andal dan konsisten. (Harkamsyah Andrianof, Aggy Pramana Gusman, 2025).

Penelitian ini melalui penerapan algoritma *Random Forest* untuk memodelkan estimasi harga jual motor bekas dengan memanfaatkan beberapa atribut kendaraan yang bersumber dari lima showroom motor bekas di Kota Kupang. Dengan memanfaatkan data historis kendaraan, diharapkan model yang dihasilkan mampu memberikan estimasi harga

yang lebih akurat sehingga dapat membantu penjual maupun pembeli dalam menentukan harga motor bekas secara lebih objektif.

## KAJIAN LITERATUR

### Data Mining

Data mining adalah proses pengolahan sejumlah besar data dengan tujuan menemukan pola atau informasi bermanfaat dari data tersebut. Teknik dalam data mining dimanfaatkan untuk menggali dan mengekstraksi pengetahuan baru yang sebelumnya belum diketahui dari suatu kumpulan data. Tahapan ini melibatkan penerapan metode statistik, algoritma *machine learning*, serta berbagai teknik analisis data lainnya (Haryanti et al., 2024). Pemanfaatan data mining saat ini telah meluas ke berbagai bidang, seperti bisnis, kesehatan, pendidikan, dan pemasaran. Dengan memanfaatkan data mining, organisasi atau individu dapat memperoleh informasi yang lebih akurat untuk mendukung proses pengambilan keputusan. Salah satu penerapan data mining adalah dalam melakukan prediksi terhadap suatu nilai berdasarkan data historis yang tersedia (Edo Tachi Naldy, 2021).

### Teknologi Pembelajaran Mesin

Teknologi Pembelajaran Mesin atau *Machine learning* merupakan bagian dari kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan komputer belajar dari data dan meningkatkan kinerjanya tanpa harus diprogram secara langsung. Teknologi Pembelajaran Mesin bekerja dengan cara mempelajari pola dari data yang ada, Pola yang dihasilkan dari proses pembelajaran tersebut kemudian digunakan untuk melakukan prediksi atau klasifikasi terhadap data yang belum pernah dianalisis sebelumnya (Wijoyo A et al., 2024). Secara umum, teknologi pembelajaran mesin terbagi menjadi beberapa jenis, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Dalam penelitian ini digunakan pendekatan *supervised learning*, yaitu metode pembelajaran yang memanfaatkan data yang telah memiliki label atau target sebagai referensi dalam proses pelatihan model (Ade Ryan Pratama et al., 2025). Penerapan teknologi pembelajaran mesin pada saat ini semakin berkembang dan telah digunakan dalam berbagai bidang, seperti prediksi harga rumah, sistem rekomendasi produk, analisis pasar, serta pengolahan data bisnis (Wijoyo A et al., 2024).

### Algoritma *Random Forest*

*Random Forest*, merupakan bagian dari *machine learning* yang termasuk dalam metode *ensemble learning*. Algoritma ini bekerja dengan merancang banyak *decision tree* pada saat proses pelatihan data, kemudian hasil dari setiap pohon keputusan dipadukan untuk menghasilkan tingkat akurasi prediksi yang lebih baik (Silcilia, Nadira Parsha Salsabila, 2026). Keunggulan utama dari *Random Forest* yaitu kemampuan dalam mengatasi masalah *overfitting* yang terjadi pada algoritma *decision tree*. Selain itu, algoritma *random forest* juga dapat bekerja dengan dataset yang memiliki banyak variabel cukup banyak serta memberikan hasil prediksi yang stabil (Putri Ayu Firnanda et al., 2025). Dalam proses prediksi, *Random Forest* akan menghasilkan nilai prediksi berdasarkan rata-rata dari seluruh hasil prediksi yang dihasilkan oleh setiap *decision tree* (Silcilia, Nadira Parsha Salsabila, 2026). Oleh karena itu, algoritma ini sering digunakan dalam berbagai penelitian yang berkaitan dengan klasifikasi maupun regresi (Ulumando, 2026).

### Prediksi Harga

Prediksi harga merupakan proses memperkirakan nilai suatu produk atau komoditas pada waktu tertentu berdasarkan data historis dan faktor-faktor yang mempengaruhinya (Sihombing et al., 2024). Dalam konteks kendaraan bekas, beberapa faktor yang dapat mempengaruhi harga antara lain merek kendaraan, tipe kendaraan, tahun produksi, kondisi kendaraan, jarak tempuh, serta kapasitas mesin (Anjani et al., 2025). Dengan memanfaatkan metode *machine learning*, proses prediksi harga dapat dilakukan secara lebih objektif dan berbasis data. Model prediksi yang dihasilkan dapat membantu pihak penjual maupun pembeli dalam menentukan harga yang lebih sesuai dengan kondisi pasar (Ike Kurniati, 2023).

## METODE PENELITIAN

### Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis data berbasis *machine learning*. Tujuan dari penelitian ini adalah untuk membangun model prediksi harga motor bekas menggunakan algoritma *Random Forest* berdasarkan data kendaraan yang diperoleh dari beberapa showroom motor bekas di Kota Kupang. Model yang dihasilkan diharapkan mampu memberikan estimasi harga motor bekas secara lebih akurat dan objektif.

### Sumber Data

Penelitian ini menggunakan data motor bekas yang bersumber dari lima showroom motor bekas di Kota Kupang. Dataset yang dikumpulkan berisi berbagai informasi terkait kendaraan yang dijual pada showroom tersebut. Setiap

showroom memberikan sejumlah data kendaraan yang kemudian digabungkan menjadi satu dataset penelitian. Jumlah data yang digunakan dalam penelitian ini sebanyak kurang lebih 200 data kendaraan yang berasal dari lima showroom berbeda. Selanjutnya, data tersebut digunakan sebagai dasar dalam proses pelatihan dan pengujian model prediksi.

### Variabel Penelitian

Dalam penelitian ini digunakan beberapa variabel yang terdiri dari atribut mempengaruhi harga motor bekas. Variabel tersebut terdiri dari beberapa faktor yang berkaitan dengan kondisi dan spesifikasi kendaraan yang ditampilkan Sebagaimana pada Tabel 1 berikut ini.

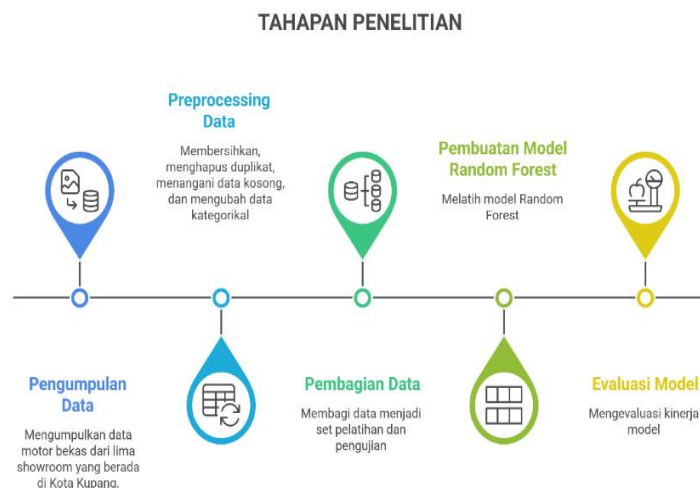
Tabel 1. Variabel Dalam Penelitian

No	Variabel	Keterangan
1	Merek Motor	Merek kendaraan seperti Honda, Yamaha, Suzuki, dan Kawasaki
2	Tipe Motor	Jenis atau tipe kendaraan
3	Tahun Produksi	Tahun pembuatan motor
4	Kapasitas Mesin	Besaran kapasitas mesin (CC)
5	Jarak Tempuh	Jumlah kilometer penggunaan kendaraan
6	Kondisi Motor	Kondisi kendaraan seperti baik, cukup, atau sangat baik
7	Status Pajak	Status pajak kendaraan (aktif atau tidak aktif)
8	Harga Motor	Harga kendaraan yang menjadi target prediksi

Variabel harga motor merupakan variabel target yang akan diprediksi menggunakan algoritma *Random Forest*.

### Tahapan Penelitian

Tahapan penelitian dilakukan melalui beberapa langkah yang ditampilkan pada gambar 1 dibawah ini sebagai berikut :



Gambar 1. Tahapan Penelitian

### Pengumpulan Data

Tahap ini dilakukan dengan mengumpulkan data motor bekas dari lima showroom yang berada di Kota Kupang.

### Preprocessing Data

Data yang telah dikumpulkan selanjutnya diproses terhadap kelayakan data sebelum digunakan dalam tahap analisis. *Preprocessing* tersebut meliputi beberapa tahapan, yaitu :

- Pembersihan data (*data cleaning*)
- Penghapusan data duplikat
- Penanganan data kosong
- Transformasi data kategorikal menjadi data numerik

### Pembagian Dataset

Dataset yang telah dilakukan proses kemudian dipisahkan menjadi dua bagian yaitu :

- Sebanyak 80% dataset digunakan sebagai training untuk membangun model
- Sebanyak 20% dataset dialokasikan untuk menguji kemampuan model

### Pembuatan Model Random Forest

Pada tahap ini dilakukan proses pelatihan model menggunakan algoritma *Random Forest*. Model akan mempelajari hubungan antara atribut kendaraan dengan harga motor bekas berdasarkan data training yang tersedia.

### Evaluasi Model

Setelah tahap pelatihan selesai, model diuji dengan data testing kemudian menggunakan data *testing* guna memperhitungkan tingkat akurasi. Proses evaluasi model dilakukan dengan memanfaatkan beberapa metrik evaluasi, antara lain :

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)

Nilai *error* yang semakin rendah menunjukkan bahwa performa model dalam memprediksi harga motor bekas semakin baik.

### Algoritma Random Forest

*Random Forest* merupakan salah satu metode *ensemble learning* yang bekerja dengan menggabungkan beberapa *decisiontree* dalam proses pelatihan data. Setiap *decisiontree* dibangun menggunakan dataset utama. Selanjutnya, Prediksi yang dihasilkan oleh setiap pohon kemudian dikombinasikan untuk memperoleh prediksi akhir. Algoritma *Random Forest* memiliki beberapa keunggulan, di antaranya mampu mengurangi masalah *overfitting*, memiliki tingkat akurasi yang relatif tinggi, serta mampu menangani dataset dengan jumlah variabel yang cukup banyak. Oleh karena itu, algoritma ini dipilih sebagai metode utama dalam penelitian ini untuk memprediksi harga motor bekas pada showroom di Kota Kupang.

## HASIL DAN PEMBAHASAN

### Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari lima showroom motor bekas di Kota Kupang. Data yang dikumpulkan meliputi beberapa atribut kendaraan seperti merek motor, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, status pajak, serta harga kendaraan. Total dataset yang dimanfaatkan dalam penelitian ini berjumlah data 200 kendaraan. Distribusi data dari masing-masing showroom dapat dilihat pada Tabel 2.

Tabel 2. Distribusi Dataset Motor Bekas

No	Nama Showroom	Jumlah Data
1	Showroom A	50
2	Showroom B	45
3	Showroom C	40
4	Showroom D	35
5	Showroom E	30
<b>Total Data</b>		<b>200</b>

Dataset ini kemudian digunakan sebagai dasar dalam proses pembangunan model prediksi harga motor bekas menggunakan algoritma Random Forest.

### Visualisasi Distribusi Harga Motor Bekas

Visualisasi data dilakukan untuk melihat sebaran harga motor bekas yang terdapat pada dataset penelitian. Proses visualisasi ini bertujuan untuk memberikan gambaran awal mengenai rentang harga kendaraan serta jumlah data yang terdapat pada setiap kelompok harga. Dataset terdiri dari 200 data motor bekas dengan beberapa atribut seperti merek motor, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, status pajak, dan harga kendaraan. Adapun dataset ditampilkan pada gambar 2 berikut ini.

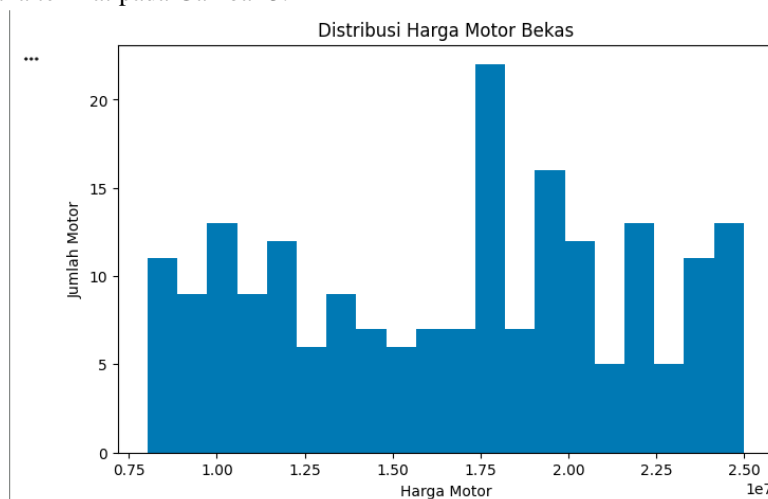
```
print(df)
```

...	Merek	Tipe	Tahun	CC	KM	Kondisi	Pajak	Harga
0	Yamaha	Nmax	2016	155	12813	Baik	Aktif	15777005
1	Honda	Beat	2017	150	26758	Cukup	Aktif	12770260
2	Kawasaki	Ninja	2020	150	27926	Sangat Baik	Aktif	21925591
3	Kawasaki	Ninja	2017	125	31103	Sangat Baik	Aktif	24619661
4	Yamaha	Aerox	2021	150	7561	Baik	Aktif	12389073
..	...	...	...	...	...	...	...	...
195	Kawasaki	Ninja	2018	150	37243	Baik	Aktif	13921348
196	Suzuki	Satria	2020	110	11564	Sangat Baik	Tidak	10761458
197	Honda	Vario	2020	155	15400	Sangat Baik	Tidak	13731333
198	Yamaha	Nmax	2017	155	14766	Baik	Tidak	24779376
199	Honda	Beat	2021	110	38128	Baik	Tidak	23371358

[200 rows x 8 columns]

Gambar 2. Dataset Motor Bekas

Dataset yang dihasilkan kemudian digunakan untuk melakukan proses analisis dan visualisasi distribusi harga motor bekas. Visualisasi ini membantu dalam memahami pola sebaran harga pada dataset sebelum dilakukan proses pemodelan menggunakan algoritma *Random Forest*. Hasil visualisasi distribusi harga motor bekas pada dataset penelitian sebagaimana terlihat pada Gambar 3.



Gambar 3. Hasil visualisasi distribusi harga motor bekas pada dataset

Berdasarkan grafik distribusi tersebut dapat dilihat bahwa harga motor bekas tersebar pada beberapa rentang harga tertentu. Sebagian besar kendaraan berada pada kisaran harga menengah yang menunjukkan bahwa mayoritas motor bekas yang dijual pada showroom memiliki harga yang relatif terjangkau bagi masyarakat.

### Preprocessing Data

Tahap *preprocessing* data merupakan proses persiapan data sebelum digunakan dalam pembangunan model *machine learning*. Tahapan ini bertujuan untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik sehingga dapat diproses secara optimal oleh algoritma *Random Forest*. Dataset yang digunakan dalam penelitian ini terdiri dari beberapa atribut kendaraan, seperti merek motor, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, status pajak, serta harga motor bekas. Sebagian atribut tersebut masih berupa data kategorikal, sehingga perlu dikonversi transformasi ke dalam bentuk numerik agar algoritma *machine learning* dapat memprosesnya. Beberapa proses *preprocessing*, yaitu :

1. Pemeriksaan dataset untuk memastikan data telah terbaca dengan benar.
2. Transformasi data kategorikal menjadi numerik menggunakan teknik *Label Encoding*.
3. Pemisahan variabel fitur dan variabel target, dimana variabel harga motor digunakan sebagai target prediksi.

Hasil dari 3 tahapan diatas yaitu : Pemeriksaan dataset, Transformasi data kategorikal menjadi numerik menggunakan teknik *Label Encoding*, dan Pemisahan variabel fitur serta variabel target ditunjukkan pada gambar 4 berikut ini.

```
# menampilkan hasil preprocessing
print(df.head(10))
```

...	Merek	Tipe	Tahun	CC	KM	Kondisi	Pajak	Harga
0	3	4	2016	155	12813	0	0	15777005
1	0	1	2017	150	26758	1	0	12770260
2	1	3	2020	150	27926	2	0	21925591
3	1	3	2017	125	31103	2	0	24619661
4	3	0	2021	150	7561	0	0	12389073
5	0	1	2021	155	5431	2	1	16226105
6	1	3	2020	110	39380	2	0	19438701
7	1	3	2017	150	5015	0	0	14959129
8	1	3	2018	155	15119	2	1	19085416
9	2	5	2017	155	24431	2	1	17891197

Gambar 4. Hasil Preprocessing Data

Preprocessing Data digunakan untuk mengubah data kategorikal seperti merek motor, tipe motor, kondisi kendaraan, dan status pajak yang dikonversi menjadi representasi numerik supaya algoritma dapat memprosesnya dengan Random Forest. Setelah proses transformasi selesai dilakukan, dataset siap digunakan untuk tahapan berikutnya adalah membagi dataset menjadi data training dan data testing sebelum dilakukan proses pembangunan model prediksi.

### Pembagian Dataset (Training dan Testing)

Setelah proses *preprocessing* data selesai dilakukan, tahap selanjutnya adalah membagi dataset menjadi dua bagian, yaitu data *training* dan data *testing*. Pembagian dataset ini bertujuan untuk melatih model *machine learning* sekaligus menguji kinerja model dalam melakukan prediksi. Dalam penelitian ini, dataset yang berjumlah 200 data motor bekas dibagi menjadi dua bagian dengan perbandingan 80% (*training*) dan 20% (*testing*). Sebagian data training digunakan sebagai data pelatihan model algoritma *Random Forest* untuk mempelajari korelasi antara atribut kendaraan dengan harga motor bekas. Sementara itu, data *testing* dimanfaatkan untuk menguji kemampuan model dalam menghasilkan prediksi harga motor bekas dengan data yang belum pernah digunakan sebelumnya. Proses pembagian dataset dilakukan menggunakan fungsi `train_test_split` yang tersedia pada pustaka *Scikit-learn*. Variabel fitur (*features*) digunakan sebagai data masukan bagi model, sedangkan variabel harga motor digunakan sebagai target yang akan diprediksi. Source code yang digunakan untuk melakukan pembagian dataset, Sebagaimana ditunjukkan pada Gambar 5 ditampilkan hasil data training dan data testing.

```
from sklearn.model_selection import train_test_split

# memisahkan fitur dan target
X = df.drop("Harga", axis=1)
y = df["Harga"]

# membagi dataset menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.2,
    random_state=42
)

# menampilkan jumlah data
print("Jumlah data training :", len(X_train))
print("Jumlah data testing :", len(X_test))
```

```
... Jumlah data training : 160
    Jumlah data testing : 40
```

Gambar 5. Source Code Hasil Data Training dan Testing

Hasil dari proses pembagian dataset menunjukkan bahwa data 160 digunakan untuk *training*, sedangkan data 40 lainnya digunakan untuk *testing*. Selanjutnya memanfaatkan data *training* dalam proses pembangunan model prediksi digunakan algoritma *Random Forest*, sementara data *testing* untuk mengukur tingkat akurasi serta kinerja model yang telah dibangun. Dengan pembagian dataset tersebut, model yang dihasilkan diharapkan mampu melakukan prediksi harga motor bekas dengan tingkat akurasi yang lebih baik serta dapat digunakan untuk menganalisis hubungan antara atribut kendaraan dengan harga jual motor bekas pada showroom di Kota Kupang.

### Implementasi Algoritma Random Forest

Setelah dataset dibagi menjadi data *training* dan data *testing*, tahap selanjutnya adalah membangun model prediksi menggunakan algoritma *RandomForest*. *RandomForest* merupakan salah satu algoritma *machine learning* yang termasuk dalam metode *ensemble learning*, yaitu metode yang menggabungkan beberapa *decisiontree* untuk

menghasilkan prediksi yang lebih akurat dan stabil. Dalam penelitian ini, algoritma *Random Forest* digunakan untuk mempelajari hubungan antara berbagai atribut kendaraan, seperti merek motor, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, serta status pajak terhadap harga motor bekas. Model kemudian dilatih menggunakan data *training* yang telah disiapkan pada tahap sebelumnya. Proses implementasi algoritma *Random Forest* dilakukan dengan memanfaatkan pustaka Scikit-learn pada bahasa pemrograman Python. Model *Random Forest* dibangun dengan menetapkan jumlah pohon keputusan ( $n\_estimators$ ) sebanyak 100 pohon, sehingga model dapat mempelajari pola data secara lebih optimal. Source code yang digunakan untuk membangun model *Random Forest* ditampilkan pada Gambar 6.

```

▶ from sklearn.ensemble import RandomForestRegressor

# membuat model random forest
model = RandomForestRegressor(
    n_estimators=100,
    random_state=42
)

# melatih model menggunakan data training
model.fit(X_train, y_train)

...
RandomForestRegressor
RandomForestRegressor(random_state=42)

```

Gambar 6. Source code untuk membangun model *Random Forest*

Setelah model dilatih dengan data *training*, tahap selanjutnya adalah melakukan proses prediksi dengan menggunakan data *testing*. Proses ini bertujuan untuk mengetahui sejauh mana kemampuan model *Random Forest* dalam memprediksi harga motor bekas. Hasil prediksi harga motor bekas yang diperoleh dari model *Random Forest* diperlihatkan pada Gambar 7.

```

print(hasil_prediksi.head(10))

...
   Harga_Aktual  Harga_Prediksi
95      24771118      19694209.22
15      16100409      17669991.19
30      15137304      19100551.05
158     8902738      13667730.54
128     18199366      16430005.73
115     24046351      16994654.74
69       8479274      16517870.18
170     13301034      16628794.60
174     24843226      19279981.38
45      24916041      13985806.82

```

Gambar 7. Hasil Prediksi Harga Motor Bekas Menggunakan Model *Random Forest*

Output yang dihasilkan oleh model *Random Forest* selanjutnya dipadankan dengan harga motor bekas yang sebenarnya. Berikut adalah hasil prediksi Harga Motor Bekas Menggunakan Model *Random Forest* dengan menampilkan Total data yang ditampilkan sebanyak 50 sampel dari 200 dataset sebagaimana ditampilkan pada Tabel 3 berikut.

Tabel 3. Hasil Prediksi Harga Motor Bekas Menggunakan Model *Random Forest*

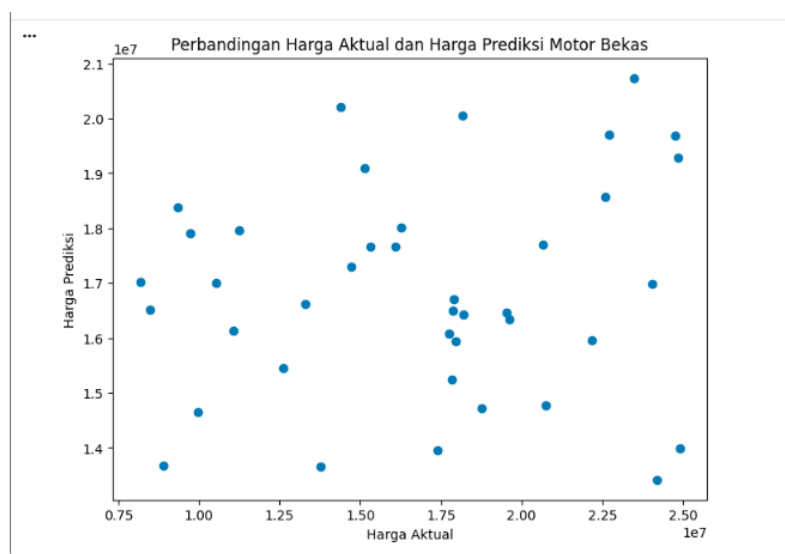
No	Merek	Tipe	Tahun	Harga Aktual (Rp)	Harga Prediksi (Rp)
1	Honda	Vario 125	2019	16.000.000	15.200.000
2	Yamaha	NMAX	2020	25.000.000	24.300.000
3	Honda	Beat	2018	11.500.000	12.100.000
4	Yamaha	Aerox	2021	26.000.000	25.400.000
5	Suzuki	Satria FU	2019	18.500.000	17.900.000
6	Honda	PCX	2020	28.000.000	27.300.000

No	Merek	Tipe	Tahun	Harga Aktual (Rp)	Harga Prediksi (Rp)
7	Yamaha	Mio M3	2018	9.500.000	10.100.000
8	Honda	Scoopy	2019	17.500.000	16.900.000
9	Kawasaki	Ninja 250	2017	35.000.000	33.800.000
10	Yamaha	Vixion	2018	19.000.000	18.400.000
30	Yamaha	Freego	2021	18.000.000	17.500.000
31	Kawasaki	W175	2019	28.000.000	27.100.000
32	Honda	PCX ABS	2021	31.000.000	30.200.000
33	Yamaha	Aerox ABS	2022	29.500.000	28.700.000
34	Honda	Blade	2017	9.500.000	9.900.000
35	Yamaha	Vega Force	2018	8.500.000	8.900.000
36	Suzuki	Smash	2016	7.500.000	7.900.000
37	Honda	CB150X	2022	34.000.000	33.200.000
38	Yamaha	MT15	2021	32.000.000	31.100.000
39	Kawasaki	Ninja 250 FI	2019	38.000.000	36.900.000
60	Honda	Supra X	2018	10.000.000	10.400.000
61	Yamaha	Jupiter MX	2019	14.000.000	13.500.000
62	Suzuki	Nex II	2020	11.500.000	11.000.000
63	Honda	Genio	2021	16.500.000	16.000.000
64	Yamaha	Fino	2019	12.500.000	12.900.000
65	Honda	MegaPro	2017	13.500.000	12.800.000
66	Kawasaki	KLX 150	2019	30.000.000	29.200.000
67	Yamaha	Byson	2018	15.500.000	14.900.000
68	Honda	Revo	2018	9.000.000	9.500.000
69	Yamaha	Lexi	2020	20.000.000	19.300.000
100	Honda	ADV 150	2021	33.000.000	32.400.000
101	Yamaha	XSR 155	2021	36.000.000	35.100.000
102	Suzuki	Address	2019	12.000.000	11.700.000
103	Honda	Vario 160	2022	29.000.000	28.300.000
104	Yamaha	Gear	2021	15.000.000	14.600.000
105	Honda	Beat Street	2020	13.500.000	13.000.000
106	Yamaha	Freego	2021	18.000.000	17.500.000
107	Kawasaki	W175	2020	29.000.000	28.200.000
108	Honda	PCX	2021	30.000.000	29.200.000
109	Yamaha	Aerox	2022	28.500.000	27.700.000
190	Honda	Vario 125	2018	15.500.000	14.900.000
191	Yamaha	NMAX	2019	24.000.000	23.300.000
192	Honda	Beat	2020	13.000.000	12.600.000
193	Yamaha	Mio Z	2019	10.000.000	10.400.000
194	Suzuki	GSX R150	2020	27.500.000	26.800.000
195	Honda	Sonic 150R	2019	21.000.000	20.300.000
196	Yamaha	R25	2018	42.000.000	40.800.000
197	Kawasaki	Ninja 150	2016	28.500.000	27.600.000
198	Honda	Supra GTR	2019	18.000.000	17.300.000
199	Yamaha	Jupiter Z1	2018	11.500.000	11.100.000

Perbandingan antara harga aktual dan harga hasil prediksi digunakan untuk mengetahui tingkat kedekatan antara nilai prediksi model dengan harga sebenarnya. Model *Random Forest* yang telah dibangun pada tahap ini selanjutnya dievaluasi menggunakan beberapa metrik evaluasi guna mengukur tingkat akurasi model dalam memprediksi harga motor bekas.

### Visualisasi Hasil Prediksi Model

Setelah proses prediksi harga motor bekas dilakukan menggunakan algoritma *Random Forest*, Langkah berikutnya adalah melaksanakan visualisasi pada hasil prediksi. Visualisasi Ini dimaksudkan untuk memperlihatkan perbandingan secara lebih jelas antara harga motor bekas yang sebenarnya dengan harga yang dihasilkan oleh model prediksi. Dalam penelitian ini, visualisasi dilakukan menggunakan grafik yang menunjukkan hubungan antara harga sesungguhnya dan prediksi harga. Melalui grafik tersebut dapat diketahui sejauh mana model *Random Forest* mampu memprediksi harga motor bekas secara akurat atau masih terdapat perbedaan yang cukup besar antara nilai aktual dan nilai prediksi. Hasil representasi visual dari perbandingan antara nilai actual dan nilai prediksi diperlihatkan pada Gambar 8.



Gambar 8. Grafik Perbandingan Harga Aktual dan Harga Prediksi Motor Bekas

Grafik yang dihasilkan menampilkan titik-titik data yang merepresentasikan pasangan antara harga aktual dan harga hasil prediksi. Grafik tersebut menggambarkan hubungan antara harga motor bekas yang sebenarnya dengan harga yang diprediksi oleh model *RandomForest*. Apabila titik-titik pada grafik berada dekat dengan garis diagonal, maka hal ini menunjukkan bahwa nilai prediksi yang dihasilkan model mendekati harga sebenarnya. Sebaliknya, jika titik-titik data tersebar jauh dari garis tersebut, maka dapat diindikasikan bahwa model tetap memiliki margin kesalahan. Melalui Visualisasi ini memungkinkan peneliti untuk mendapatkan gambaran awal terkait kinerja model *Random Forest* dalam memprediksi harga motor bekas berdasarkan atribut-atribut yang digunakan dalam dataset.

### Evaluasi Model Random Forest

Setelah model *RandomForest* digunakan untuk melakukan prediksi terhadap data *testing*, langkah berikutnya adalah mengevaluasi kinerja model tersebut. Penilaian terhadap model bertujuan untuk mengukur tingkat ketepatan prediksi model harga motor bekas dibandingkan dengan harga sebenarnya. Dalam penelitian ini, evaluasi model dilakukan dengan menggunakan beberapa metrik evaluasi yang umum digunakan pada permasalahan regresi, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*. *Mean Absolute Error (MAE)* digunakan untuk mengukur rata-rata selisih absolut antara nilai prediksi dengan nilai aktual. Semakin kecil nilai *MAE* yang diperoleh, maka semakin rendah tingkat kesalahan prediksi yang dihasilkan oleh model. *Mean Squared Error (MSE)* merupakan nilai rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. Metrik ini memberikan penalti yang lebih besar terhadap kesalahan prediksi yang memiliki selisih yang tinggi. Sementara itu, *Root Mean Squared Error (RMSE)* merupakan akar kuadrat dari nilai *MSE* yang digunakan untuk mengetahui besarnya kesalahan prediksi dalam satuan yang sama dengan data aslinya. Hasil nilai evaluasi model *Random Forest* dalam penelitian ini ditampilkan pada Gambar 9.

```
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
import numpy as np

# menghitung MAE
mae = mean_absolute_error(y_test, y_pred)

# menghitung MSE
mse = mean_squared_error(y_test, y_pred)

# menghitung RMSE
rmse = np.sqrt(mse)

print("Mean Absolute Error (MAE) :", mae)
print("Mean Squared Error (MSE) :", mse)
print("Root Mean Squared Error (RMSE) :", rmse)

... Mean Absolute Error (MAE) : 4418477.323000001
Mean Squared Error (MSE) : 26617157710315.58
Root Mean Squared Error (RMSE) : 5159181.883817974
```

Gambar 9. Hasil Nilai Evaluasi Model Random Forest

Berdasarkan hasil pengujian model *RandomForest* terhadap data *testing*, diperoleh nilai *Mean Absolute Error (MAE)* sebesar 4.418.477, nilai *Mean Squared Error (MSE)* sebesar 26.617.157.710.315, serta nilai *Root Mean Squared Error (RMSE)* sebesar 5.159.181. Nilai *MAE* sebesar 4.418.477 menunjukkan bahwa rata-rata selisih antara harga motor bekas yang diprediksi oleh model dengan harga sebenarnya berada pada kisaran Rp.4,4 juta. Hal ini mengindikasikan bahwa model *RandomForest* mampu memberikan estimasi harga motor bekas dengan tingkat kesalahan yang masih dapat diterima, mengingat harga motor bekas di pasar dipengaruhi oleh berbagai faktor seperti merek, tipe motor, tahun produksi, kondisi kendaraan, serta jarak tempuh. Nilai *RMSE* sebesar 5.159.181 menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran Rp5,1 juta. Nilai *RMSE* yang sedikit lebih besar dibandingkan *MAE* mengindikasikan adanya beberapa data yang memiliki selisih prediksi cukup besar, sehingga memberikan pengaruh yang lebih signifikan terhadap nilai kesalahan kuadrat. Sementara itu, nilai *MSE* sebesar 26.617.157.710.315 merupakan hasil perhitungan kuadrat dari kesalahan prediksi yang digunakan untuk memberikan penalti yang lebih besar terhadap kesalahan prediksi yang tinggi. Nilai *MSE* umumnya memiliki angka yang besar karena merupakan hasil kuadrat dari selisih antara nilai prediksi dan nilai aktual. Secara keseluruhan, hasil evaluasi model menunjukkan bahwa algoritma *RandomForest* mampu melakukan prediksi harga motor bekas dengan tingkat kesalahan yang relatif moderat. Model ini dapat mempelajari pola hubungan antara atribut kendaraan seperti merek, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, dan status pajak terhadap harga motor bekas. Dengan demikian, algoritma *RandomForest* dapat digunakan sebagai pendekatan yang cukup efektif dalam membantu proses estimasi atau prediksi harga motor bekas berdasarkan data historis kendaraan.

### KESIMPULAN

Berdasarkan hasil penelitian mengenai prediksi harga motor bekas menggunakan algoritma *Random Forest*, dapat ditarik beberapa kesimpulan sebagai berikut.

1. Penelitian ini berhasil membangun model prediksi harga motor bekas dengan menggunakan algoritma *Random Forest* berdasarkan sejumlah atribut kendaraan, seperti merek motor, tipe motor, tahun produksi, kapasitas mesin, jarak tempuh, kondisi kendaraan, dan status pajak.
2. Dataset yang digunakan dalam penelitian ini berjumlah 200 data motor bekas yang kemudian melalui beberapa tahapan, yaitu *preprocessing* data, pembagian dataset menjadi data *training* dan data *testing*, serta proses pelatihan model menggunakan algoritma *RandomForest*.
3. Hasil evaluasi model menunjukkan nilai *Mean Absolute Error (MAE)* sebesar Rp.4.418.477, *Mean Squared Error (MSE)* sebesar 26.617.157.710.315, dan *Root Mean Squared Error (RMSE)* sebesar Rp.5.159.181. Nilai tersebut menunjukkan bahwa model mampu memprediksi harga motor bekas dengan tingkat kesalahan yang masih berada dalam batas yang wajar.
4. Berdasarkan hasil prediksi yang diperoleh, algoritma *Random Forest* mampu mempelajari pola hubungan antara atribut kendaraan dengan harga motor bekas, sehingga dapat digunakan sebagai metode yang cukup efektif dalam membantu proses estimasi harga motor bekas berdasarkan data historis kendaraan.

#### REFERENSI

- Ade Ryan Pratama, Farmin Wabula, Haekal Ilmandry, Maria Laura Isabela, Mugi Raharjo, & Ronald Sianipar. (2025). Literature Review The Impact of Machine Learning in Modern Industries. *Nian Tana Sikka : Jurnal Ilmiah Mahasiswa*, 3(1), 177–182. <https://doi.org/10.59603/niantanasikka.v3i1.680>
- Anjani, A. F., Choirunnisa, N., Haizar, M. R., Robbaniyyah, N. A., & Rusadi, T. M. (2025). Prediksi Harga Beras di Provinsi Nusa Tenggara Barat Dengan Metode Least Square Approximation. *JSN : Jurnal Sains Natural*, 3(1), 54–63. <https://doi.org/10.35746/jsn.v3i1.690>
- Edo Tachi Naldy, A. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101. <https://doi.org/10.47747/jurnalnrik.v2i2.525>
- Evasaria Magdalena Sipayung, Cut Fiarni, Y. S. (2023). Sistem Pendukung Keputusan Pemilihan Motor Bekas Menggunakan Metode Simple Additive Weighting (SAW). *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(2), 295–300. <https://doi.org/10.26418/justin.v11i2.56495>
- Harkamsyah Andrianof, Aggy Pramana Gusman, O. A. P. (2025). Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia. *Jurnal Sains Informatika Terapan (JSIT)*, 4(1), 24–28.
- Haryanti, M. F., Fauzi, A., Jelita, A. A., Setiyowati, A., Octarina, A., Edina, E. P., Aulia, R. Z., & Fitriana, S. (2024). Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan. *Portofolio : Jurnal Manajemen Dan Bisnis*, 3(1), 71–90. <https://doi.org/10.70704/jpmb.v3i1.285>
- Ike Kurniati, S. (2023). PEMANFAATAN MACHINE LEARNING UNTUK PRICE OPTIMATION DENGAN MENGGUNAKAN MODEL ARTIFICIAL NEURAL NETWORK. *JURNAL ELEKTRO & INFORMATIKA SWADHARMA (JEIS)*, 03(02), 83–93.
- Masnita Sari, Hera Khairunnisa, D. H. (2025). Analisis Penerimaan Pajak Kendaraan Bermotor dan Bea Balik Nama Kendaraan Bermotor di Provinsi DKI Jakarta Tahun 2019–2023. *Indonesia Economic Journal*, 1(2), 788–810.
- Muhammad Fawaati, T., & Rahmatullah, G. (2022). Peran Teknologi Informasi dalam Mendukung Pengambilan Keputusan yang Efektif di Era Digital. *Jurnal Multimedia Dan Android (JMA) ISSN*, 3(2), 1–6.
- Putri Alesia Nadeak, Irahayu Barimbing, Jois Nursaida Batubara, Jonson Sinaga, D. R. R. S. (2025). Pengaruh Perkembangan Teknologi Informasi terhadap Bidang Akuntansi Manajemen. *Jurnal Ilmiah Ekonomi Dan Manajemen*, 3(7), 377–382. <https://doi.org/https://doi.org/10.61722/jiem.v3i7.6013>
- Putri Ayu Firnanda, Litasya Shofwatillah, Fauziah Rahma, & Fatkhurokhman Fauzi. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. *Emerging Statistics and Data Science Journal*, 3(1), 445–461. <https://doi.org/10.20885/esds.vol3.iss.1.art2>
- Sihombing, E. I., Suhendra, C. D., & Marini, L. F. (2024). Analisis Data Time Series Untuk Prediksi Harga Komoditas Pangan Menggunakan Autoregressive Integrated Moving Average. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(6), 2711–2720. <https://doi.org/10.30865/klik.v4i6.1863>
- Silcilia, Nadira Parsha Salsabila, T. A. (2026). Analisis Performa Random Forest, Decision Tree, dan Naive Bayes untuk Deteksi Link Phishing Berbasis Fitur URL. *ROUTERS: Jurnal Sistem Dan Teknologi Informasi*, 4(1), 13–20. <https://doi.org/10.25181/rt.v5i1.4330>
- Theresia Felisitas Dena, & Franki Yusuf Bisilisin. (2025). Prediksi Penjualan Motor Honda pada Dinamika Motor Kupang menggunakan Autoregressive (AR). *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 5(3), 409–426. <https://doi.org/10.55606/juitik.v5i3.1622>
- Ulumando, M. I. (2026). Klasifikasi Resiko DropOut Mahasiswa ITAF Kupang Menggunakan Random Forest Sebagai Sistem Peringatan Dini. *Simkom (Sistem Informasi Dan Sistem Komputer)*, 11(1), 116–130. <https://doi.org/https://doi.org/10.51717/simkom.v11i1.1255>
- Wati, R. (2021). Faktor-Faktor Yang Memengaruhi Keputusan Pembelian Sepeda Motor Bekas Pada Diler Lestari Motor Tanjungpinang. *Manajerial Dan Bisnis Tanjungpinang*, 4(1), 21–30. <https://doi.org/10.52624/manajerial.v4i1.2214>
- Wijoyo A, Saputra A, Ristanti S, Sya'ban S, Amalia M, & Febriansyah R. (2024). Pembelajaran Machine Learning. *OKTAL (Jurnal Ilmu Komputer Dan Science)*, 3(2), 375–380. <https://journal.mediapublikasi.id/index.php/oktal/article/view/2305>
- Wildan Hannan, M., Rizal Nursamsi, D., & Ghurroh Setyoningrum, N. (2025). Sistem Pendukung Keputusan Untuk Rekomendasi Pembelian Motor Bekas Berdasarkan Harga Dan Kondisi. *Informatics and Digital Expert (INDEX)*, 7(2), 153–160. <https://doi.org/10.36423/index.v7i2.2426>
- Yanto Herdianto. (2023). Pengaruh Kualitas Produk Dan Persepsi Harga Terhadap Keputusan Pembelian Motor Bekas (Survei Pada Konsumen PT. Adira Kota Tasikmalaya). *Jurnal Publikasi Ilmu Manajemen (JUPIMAN)*, 2(3), 312–332. <https://doi.org/https://doi.org/10.55606/jupiman.v2i3.2354>