

Perbandingan Performa Algoritma Linear Regresi dan *Random Forest* untuk Prediksi Harga Bawang Merah di Kota Samarinda

Muhammad Aditya Pratama*¹
Mukminatul Munawaroh²
Wawan Joko Pranoto³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur, Indonesia

*e-mail: 2011102441135@umkt.ac.id¹, 20111024411064@umkt.ac.id², wjp337@umk.ac.id³

Abstrak

Sebagai komponen penting dari perekonomian, pertanian membutuhkan prediksi harga yang akurat untuk memberi tahu pemerintah dan pelaku industri. Fokus penelitian ini adalah untuk membandingkan kinerja dua algoritma mesin learning, Linear Regresi dan *Random Forest*, dalam memprediksi harga bawang merah di Kota Samarinda dengan mempertimbangkan variabel harga pasar, nilai rata-rata, nilai rata-rata sebelumnya, dan tingkat inflasi. Dataset historis dikumpulkan, yang mencakup harga pasar bawang merah di berbagai lokasi di Kota Samarinda. Sementara *Random Forest* mengumpulkan pola kompleks dan interaksi dalam data, linear regresi mengevaluasi hubungan linear antara variabel input dan output. Hasil penelitian menunjukkan bahwa Linear Regresi membuat prediksi yang lebih akurat dalam situasi hubungan linear antara variabel input dan output adalah fluktuasi harga yang dipengaruhi oleh faktor pasar dan tingkat inflasi. Dengan mempertimbangkan pasar, nilai rata-rata, nilai rata-rata sebelumnya, dan tingkat inflasi Kota Samarinda, hasil RMSE Linear Regresi adalah 53.74842694081432 dan hasil RMSE *Random Forest* adalah 841.1322301347325. Studi ini memberikan pemahaman lebih mendalam tentang bagaimana kinerja algoritma mesin learning dapat dipengaruhi oleh variasi faktor saat meramalkan harga bawang merah. Penelitian ini dapat membantu para pemangku kepentingan mengelola dan mengantisipasi perubahan harga bawang merah di pasar lokal Samarinda.

Kata kunci: Bawang Merah, Harga, Linear Regresi, Prediksi, *Random Forest*

Abstract

As an important component of the economy, agriculture requires accurate price predictions to inform governments and industry players. The focus of this research is to compare the performance of two machine learning algorithms, Linear Regression and *Random Forest*, in predicting shallot prices in Samarinda City by considering market price variables, average value, previous average value, and inflation rate. A historical dataset was collected, which includes the market price of shallots in various locations in Samarinda City. While *Random Forest* collects complex patterns and interactions in data, linear regression evaluates the linear relationship between input and output variables. The research results show that Linear Regression makes more accurate predictions in situations where the linear relationship between input and output variables is price fluctuations which are influenced by market factors and the inflation rate. By considering the market, average value, previous average value, and inflation rate for Samarinda City, the RMSE Linear Regression result is 53.74842694081432 and the RMSE *Random Forest* result is 841.1322301347325. This study provides a deeper understanding of how the performance of machine learning algorithms can be influenced by a variety of factors when forecasting onion prices. This research can help stakeholders manage and anticipate changes in shallot prices in the Samarinda local market.

Keywords: Shallots, Price, Linear Regression, Prediction, *Random Forest*

PENDAHULUAN

Bawang merah merupakan komoditas hortikultura yang memiliki nilai ekonomis yang tinggi dan peluang pasar yang besar sebagai bumbu untuk konsumsi rumah tangga, bahan baku industri serta untuk memenuhi kebutuhan ekspor. Dalam beberapa tahun terakhir harga bawang merah dalam negeri meningkat akibat adanya kesenjangan antara produksi dengan permintaan dalam negeri selain itu juga karena musim panen (tanam) bawang merah di Indonesia yang tidak

menentu diakibatkan oleh gagal panen dan juga terjadinya EL NINO atau bencana kekeringan dalam kurun waktu yang cukup lama (Alfin *et al.*, 2022).

Langkah yang dapat dilakukan untuk mengatasi permasalahan tersebut yaitu dengan membuat sistem prediksi harga Bawang Merah harian. Beberapa metode yang dapat digunakan untuk memprediksi harga yang bernilai numerik kontinu yaitu dengan menggunakan metode regresi linear dan random forest. Metode Linear Regresi berfungsi untuk memodelkan data. Proses prediksi dalam machine learning yaitu dengan memahami sifat atau ciri dari objek yang tidak dikenal dengan cara mengidentifikasi pola (pattern) dalam himpunan data (dataset) (Puteri and Silvanie, 2020).

Regresi linear merupakan metode yang dapat digunakan untuk mengukur minimal 2 variabel, cara mengukur datanya dengan menggunakan variabel dependen dan variabel independen yang digambar melalui Korelasi antar 2 variabel tersebut melalui garis lurus (Suryanto, 2019).

Random Forest Regression adalah salah satu algoritma pembelajaran machine learning yang paling populer dan sering digunakan dalam penganalisisan data. Metode ini menggunakan pendekatan supervised learning. Pendekatan ini melibatkan penerapan ide pengawasan pada set data pelatihan untuk memungkinkan proses pembangunan kelas classifier. Dengan kata lain, data berlabel diperlukan untuk model ini untuk memahami dan memprediksi pola. Pada dasarnya, model Random Forest Regression dibuat dengan menggabungkan beberapa Pohon Keputusan, di mana setiap pohon (atau pohon) bertindak sebagai pengambil keputusan independen. Algoritma ini unik karena dapat menggabungkan prediksi dari semua pohon untuk membuat prediksi akhir yang lebih akurat dan dapat diandalkan (Fitri, 2023).

Pemilihan regresi linear dan random forest sebagai metode prediksi dalam penelitian ini didasarkan pada parameter model sederhana dan keunggulan estimasi data berbasis time series. Selain itu, metode ini memungkinkan untuk melakukan analisis menggunakan beberapa variabel independen (X), yang dapat menghasilkan prediksi yang lebih akurat (Arinal and Azhari, 2023).

Berdasarkan penelitian terdahulu oleh M Fadhil Arfa, Muhammad Ridho Al Fathan, Habibah Br Lumbantobing, Rahmaddeni, pada tahun 2023 yang berjudul "Prediksi Harga Cryptocurrency dengan Metode Linear Regresi". Dalam penelitian ini dilakukan pengaplikasian metode linear regresi, pertama-tama dilakukan pengumpulan data harga cryptocurrency, volume transaksi, dan indeks harga pasar selama beberapa periode. Kemudian, data tersebut diolah dengan menggunakan software statistik seperti SPSS, Excel atau Python. Hasil dari penelitian ini adalah model linear regresi yang dapat memprediksi harga cryptocurrency dengan akurasi sebesar 91,46%. Dengan demikian, metode linear regresi dapat digunakan sebagai alternatif untuk memprediksi harga cryptocurrency (Arfa *et al.*, 2023).

Penelitian terkait dalam Penerapan metode Regresi Linear Untuk Prediksi Harga Beras di Indonesia. Mengetahui hasil prediksi harga beras dan Mengetahui nilai RMSE Penelitian ini dilakukan dengan menggunakan data harga beras pada tahun 2021 hingga tahun 2023. Hasil prediksi Harga Beras yang diperoleh dari penerapan Regresi Linear bahwa nilai prediksi harga beras mendekati nilai aktual harga beras. kemudian peneliti membagi data 70% untuk data training dan 30% untuk data testing. Dari kondisi pembagian data tersebut memperoleh nilai RMSE (Root Mean Squared Error) 337.996 +/-0.000. Pengujian menggunakan perhitungan RMSE digunakan untuk mengetahui keakuratan hasil prediksi harga beras (Arinal and Azhari, 2023).

Penelitian terkait dalam Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear di Nolbas SVNR. Penelitian ini menggunakan metode regresi linear untuk mengolah data penjualan menggunakan atribut seperti tahun, nama pelanggan, harga barang, harga bahan dan jumlah pesanan. Hasil pengujian didapatkan konstanta regresi nilai positif sebesar 179.4432, nilai koefisien B1 adalah -2.4839 dan nilai koefisien B2 adalah sebesar 0.5601. Hasil nilai konstanta dan koefisien regresi tersebut digunakan untuk melakukan prediksi jumlah produksi sablon tahun 2023 di Nolbas SVNR dan nilai prediksinya adalah 3391. Evaluasi dari model regresi linear berganda di dapatkan nilai MAE sebesar 3.7247, nilai MSE sebesar 17.8633 dan nilai r² score didapatkan sebesar 87% (Fadhilah and Ali, 2023).

Penelitian terkait dalam Penerapan Algoritma Regresi Linear Sederhana Untuk Prediksi Pengaruh Nilai Omset Terhadap Besaran Pendapatan Bersih Usaha Mikro Kecil Dan Menengah.

Penelitian dilakukan dengan memeriksa data yang salah atau dapat mengganggu proses analisis, menganalisis kumpulan data, dan membaginya menjadi data pelatihan dan data uji. Proses pembagian data adalah 80% digunakan untuk data training dan 20% untuk data test. Hasil penelitian menghasilkan model regresi linear sederhana $y = -374359.94312876 + (0.20259073(X))$ dengan nilai Correlation: 0.9999546662293942 Coefficient of determination 0.7244150380582826, Mean Absolute Error: 243269.23966642626, Mean Squared Error: 71893932500.72606 dan Root Mean Squared Error: 268130.43934012053. Hasil akhir analisis dilakukan perbandingan antara y aktual dengan y prediksi baik dalam bentuk tabel maupun grafik (Musthofa *et al.*, 2023).

Penelitian terkait dalam Implementasi Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras Di Kota Medan. Penelitian ini menghasilkan sistem yang dapat memprediksi harga beras di Kota Medan dengan menggunakan Metode Regresi Linear Sederhana. Hasil penelitian menunjukkan bahwa nilai pengukuran error metode RMSE yang rendah, yaitu 0,559861633, menunjukkan bahwa hasil prediksi metode RMSE sangat akurat. (Ong, 2009).

Penelitian terkait dalam Perbandingan performa Algoritma Neural Network, Regresi Linear, dan Random Forest dalam simulasi prediksi angka kematian pasien COVID-19 di Indonesia. Penelitian ini dirancang untuk memodelkan algoritma dan mengetahui algoritma apa yang paling cocok/mendekati nilai aslinya dalam memprediksi angka kematian berdasarkan dataset WHO dan dataset Johns Hopkins University Center for Systems Science and Engineering. Hasil dari pemodelan tersebut Pada hipotesis awal algoritma yang diduga akan paling mendekati nilai asli adalah algoritma neural network, namun setelah dilakukan perbandingan performa ternyata algoritma random forest adalah algoritma yang paling mendekati nilai asli dalam melakukan prediksi dengan nilai RMS 94.991 +/- 0.000 selanjutnya diikuti oleh algoritma neural network dengan nilai RMS 285.956 +/- 0.000 simpangan terjauh terjadi pada algoritma linear regression yaitu positive 3285 menyimpang dari nilai sebenarnya (Redo *et al.*, 2021).

Tujuan adanya penelitian ini adalah untuk melakukan prediksi harga Bawang Merah. Dengan cara melakukan penerapan metode regresi linear dan random forest. Untuk pengujian keakurasian digunakan metode RSME yang artinya hasil tersebut menunjukkan hasil yang akurat, Hal Ini menunjukkan bahwa squared Error menghasilkan nilai yang lebih tinggi dari Root Mean Squared Error (Arinal and Azhari, 2023).

Oleh karena itu, dalam penelitian ini peneliti berharap untuk menemukan algoritma terbaik untuk memprediksi harga bawang merah dengan menggunakan teknik data mining yang menggunakan algoritma Regresi Linear dan Random Forest. Penelitian ini juga berharap untuk dapat membantu pembeli bawang merah meminimalkan resiko lonjakan harga dan memaksimalkan keuntungan yang dapat mereka peroleh, serta untuk mendapat algoritma terbaik untuk prediksi harga bawang merah (Supriyanto, Ilhamsyah and Enri, 2022).

METODE

Jenis Penelitian

Penelitian ini menggunakan metode penelitian terapan. Penelitian jenis ini bertujuan untuk memecahkan masalah yang relevan, baik masalah yang dihadapi oleh individu atau kelompok. Hasil penelitian ini dapat digunakan secara langsung, seperti halnya penelitian dan pengembangan yang dilakukan untuk mengembangkan produk baru yang diuji secara empiris. Dengan kata lain, penelitian terapan bertujuan untuk menemukan jawaban atas pertanyaan tertentu. (Ong, 2009).

Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan disaat melakukan PKL (Praktik Kerja Lapangan) di sebuah instansi pemerintahan yang ada di Kota Samarinda yang dimulai dari bulan Juli 2023 sampai bulan September tahun 2023.

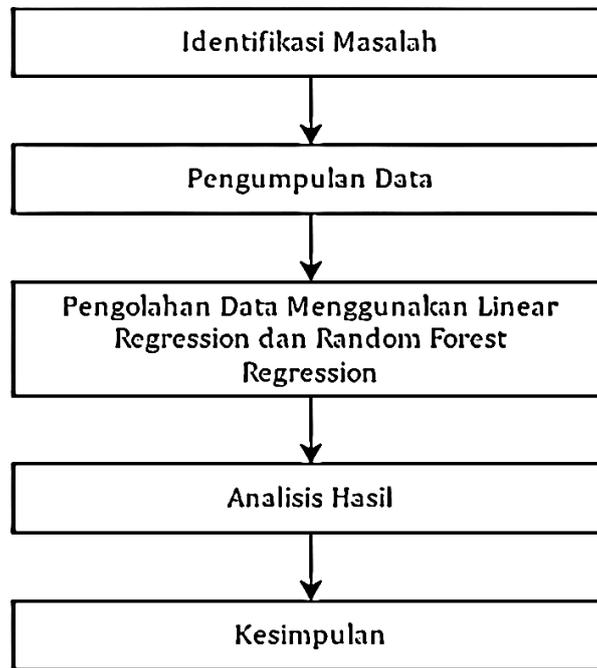
Dataset

Dataset yang digunakan dalam proses penelitian ini diolah dari Dinas Perdagangan dan UMKM Kota Samarinda, harga bawang.csv berisi 152 data dan terdiri dari 4 kolom. Ke empat kolom tersebut terdiri dari 1 kolom kategori (Tanggal, dan 3 kolom numerik (X) yaitu tanggal,

bawang merah, bawang putih dan (Y) rata-rata.

Metodologi Penelitian

Beberapa langkah yang diambil dalam penelitian ini adalah sebagai berikut:



Gambar 1. Bagan Metodologi Penelitian

1. Identifikasi Masalah

Menentukan atau identifikasi permasalahan yang akan dipecahkan, yaitu memprediksi jumlah transaksi ojek online menggunakan regresi linear dan random forest.

2. Pengumpulan Data

Langkah selanjutnya adalah melakukan pencarian dan pemilihan dataset yang akan digunakan. Dataset yang digunakan dalam proses penelitian ini diolah dari Dinas Perdagangan dan UMKM Kota Samarinda, harga bawang.csv berisi 152 data dan terdiri dari 4 kolom. Ke empat kolom tersebut terdiri dari 1 kolom kategori (Tanggal, dan 3 kolom numerik (X) yaitu tanggal, bawang merah, bawang putih dan (Y) rata-rata.

3. Pengolahan Data Menggunakan Regresi Linear dan Random Forest

Setelah dataset terkumpul, tahap berikutnya adalah melakukan pengolahan data menggunakan regresi linear dan random forest. Proses ini memiliki beberapa tahap, seperti data cleaning (penanganan outlier, transformasi log, normalisasi data), eksplorasi data (menganalisis dan memahami data yang ada), data preprocessing (pemisahan data training dan data testing), model devining (mendefinisikan model regresi linear dan random forest yang akan digunakan), dan model evaluation (evaluasi kinerja model yang telah dilatih).

4. Analisis Hasil

Setelah model dibentuk dan dievaluasi, tahap selanjutnya adalah menganalisis perbandingan hasil prediksi harga bawang merah menggunakan regresi linear dan random forest.

5. Kesimpulan

Kesimpulan diambil dari analisis kinerja metode regresi linear dan random forest dalam memprediksi harga bawang merah. Evaluasi dilakukan berdasarkan keakuratan prediksi, kegunaan metode, dan saran untuk penelitian selanjutnya.

Root Mean Squared Error

Untuk menghitung kuadrat error dan perbedaan antara nilai hasil prediksi dan nilai sebenarnya, banyaknya waktu data peramalan dibagi dengan jumlah selisih, dan kemudian menarik akar MSE. Rumus untuk menghitung nilai RMSE:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}}$$

Keterangan:

RMSE: *Mean Square Error*

Y_t : Nilai aktual pada periode t

\hat{Y}_t : Nilai peramalan pada periode t

n: Banyaknya data

HASIL DAN PEMBAHASAN

Perbandingan prediksi harga bawang merah di kota samarinda menggunakan algoritma prediksi linear regresi dan random forest berdasarkan nilai rata-rata harga dengan data sampel yang mencakup harga bawang merah dari 11 pasar di kota samarinda pada tanggal 01 januari 2021 sampai tanggal 01 januari 2023 diperoleh melalui akses ke situs web Lamin Etam. Dengan data tersebut melakukan perbandingan harga guna memprediksikan harga bawang merah pada tahun 2024 mendatang. Proses pengolahan data menggunakan perangkat lunak Python yang menerapkan algoritma linear regresi dan random forest, berikut langkah-langkahnya:

1. Import Library yang akan di gunakan

```
import pandas as pd
import seaborn as sns
%matplotlib inline
import matplotlib.pyplot as plt
import math

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from math import sqrt
from sklearn import metrics
```

Gambar 2. Library Python

Mengimpor pustaka pandas dengan alias `pd`. Pandas digunakan untuk manipulasi dan analisis data tabular. Alias `pd` digunakan untuk mempermudah pemanggilan fungsi dari pustaka. Mengimpor pustaka seaborn dengan alias `sns`. Seaborn adalah pustaka visualisasi data berdasarkan Matplotlib, dan sering digunakan untuk membuat plot yang lebih menarik dan informatif, yang memungkinkan hasil plot dari Matplotlib ditampilkan langsung di dalam notebook Matplotlib digunakan untuk membuat plot dan visualisasi data. Mengimpor modul `math`, yang menyediakan fungsi-fungsi matematika umum. Mengimpor fungsi `train_test_split` digunakan untuk membagi data menjadi data latih dan data uji untuk melatih dan menguji model.

Mengimpor kelas `LinearRegression` dari scikit-learn. Kelas ini digunakan untuk membuat model.

Mengimpor kelas `RandomForestRegressor` dari scikit-learn. Kelas ini digunakan untuk membuat model regresi menggunakan algoritma Random Forest. Mengimpor fungsi `sqrt` (akar kuadrat) dari modul `math`, digunakan untuk menghitung nilai akar kuadrat, yang dapat berguna dalam evaluasi model. Mengimpor modul `metrics` dari scikit-learn. Modul ini menyediakan berbagai metrik evaluasi yang dapat digunakan untuk mengevaluasi

performa model, seperti mean squared error (MSE) dan coefficient of determination (R-squared).

2. Import dataset dan Membaca data

```
# Baca dataset
data = pd.read_excel('Harga Bawang.xlsx')
data.head()
```

	DATE	PAGI	SUNGAI DAMA	KEMUNING	BAQA	PALARAN	SEGIRI	MERDEKA	IJABAH	KEDONDONG	LOK BAHU	BENGGURUNG	RATA-RATA	RATA-RATA SEBELUMNYA	TINGKAT INFLANSI
0	2021-01-01	28350	26200	28250	27700	23550	26200	24200	28000	27400	28600	29300	27100	27100	0.000000
1	2021-02-01	28650	26800	25900	26400	23850	27050	25150	27350	26550	27800	28150	26700	27100	-0.014760
2	2021-03-01	30850	30850	28650	29200	28250	30000	28500	30900	30000	30350	32350	30000	26700	0.123596
3	2021-04-01	29350	26500	27800	27850	26600	28300	22700	28450	28450	28500	29050	27600	30000	-0.080000
4	2021-05-01	28400	25800	27200	26700	24600	28000	23850	27950	28150	28150	28450	27000	27600	-0.021739

Gambar 3. Hasil membaca 5 data teratas

Membaca dataset dari file Excel dengan nama file 'Harga Bawang.xlsx' menggunakan pandas. Fungsi `read_excel` dari pandas digunakan untuk membaca data dari file Excel ke dalam DataFrame, yang merupakan struktur data tabular pandas. Mencetak lima baris pertama menggunakan metode `head()`. Metode ini memberikan preview atau gambaran singkat dari data yang telah dibaca. Ini membantu dalam melihat struktur dan isi data dengan cepat.

3. Processing data

```
data = data.set_index('DATE')
y = data['RATA-RATA']
X = data.drop(['RATA-RATA'], axis=1)

# Pisahkan data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

y.head()
X.head()
```

	DATE	PAGI	SUNGAI DAMA	KEMUNING	BAQA	PALARAN	SEGIRI	MERDEKA	IJABAH	KEDONDONG	LOK BAHU	BENGGURUNG	RATA-RATA	RATA-RATA SEBELUMNYA	TINGKAT INFLANSI
0	2021-01-01	28350	26200	28250	27700	23550	26200	24200	28000	27400	28600	29300	27100	27100	0.000000
1	2021-02-01	28650	26800	25900	26400	23850	27050	25150	27350	26550	27800	28150	26700	27100	-0.014760
2	2021-03-01	30850	30850	28650	29200	28250	30000	28500	30900	30000	30350	32350	30000	26700	0.123596
3	2021-04-01	29350	26500	27800	27850	26600	28300	22700	28450	28450	28500	29050	27600	30000	-0.080000
4	2021-05-01	28400	25800	27200	26700	24600	28000	23850	27950	28150	28150	28450	27000	27600	-0.021739

Menggunakan metode `set_index('DATE')` mengubah kolom atribut 'DATE' dijadikan indeks (index). Mendefinisikan variabel `y` sebagai kolom 'RATA-RATA'. Ini menunjukkan bahwa 'RATA-RATA' akan menjadi variabel dependen yang ingin diprediksi. Mendefinisikan variabel `X` sebagai DataFrame yang dihasilkan dari menghapus kolom 'RATA-RATA'. Ini menunjukkan bahwa 'RATA-RATA' akan dihapus dari variabel independen. Menggunakan `train_test_split` dari scikit-learn untuk membagi data menjadi data latih dan data uji. `X_train` dan `y_train` akan menjadi data latih untuk fit model. `X_test` dan `y_test` akan menjadi data uji untuk mengevaluasi model. Argumen `test_size=0.2` Menunjukkan bahwa 20% dari data akan digunakan sebagai data uji, sedangkan 80% akan menjadi data latih. Argumen `random_state=42` Digunakan untuk memastikan reproduktibilitas, untuk mendapatkan pembagian yang sama menjalankan skrip.

4. Implementasi Model Regresi Linear

```
# Inisialisasi model regresi linier
linreg_model = LinearRegression()

# Latih model regresi linier
linreg_model.fit(X_train, y_train)

# Prediksi menggunakan data uji
y_pred_linreg = linreg_model.predict(X_test)

from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error

# Evaluasi model Linear Regression
rmse_linreg = sqrt(mean_squared_error(y_test, y_pred_linreg))
mae_linreg = mean_absolute_error(y_test, y_pred_linreg)
mse_linreg = mean_squared_error(y_test, y_pred_linreg)

print("Linear Regression Metrics:")
print(f"RMSE: {rmse_linreg}")
print(f"MAE: {mae_linreg}")
print(f"MSE: {mse_linreg}")
```

```
Linear Regression Metrics:
RMSE: 53.74842694081432
MAE: 42.78741987216355
MSE: 2888.8933986120546
```

Berikut adalah penjelasannya:

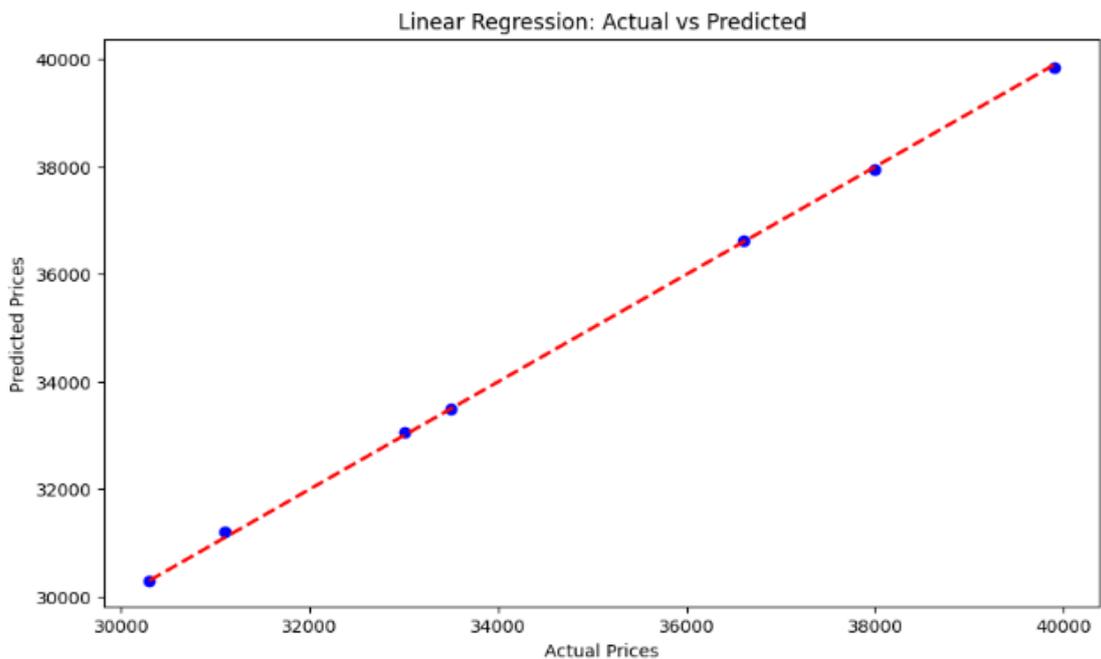
- a. Inisialisasi Model Regresi Linear
Membuat objek model regresi linear menggunakan kelas `LinearRegression` dari `scikit-learn`.
- b. Latih Model Regresi Linear
Melatih model regresi linear menggunakan data latih (`X_train` dan `y_train`). Model belajar dari hubungan antara variabel independen (`X_train`) dan variabel dependen (`y_train`).
- c. Prediksi Menggunakan Data Uji
Menggunakan model regresi linear yang telah dilatih untuk membuat prediksi terhadap data uji (`X_test`). Hasil prediksi disimpan dalam variabel `y_pred_linreg`.
- d. Evaluasi Model Linear Regression
Menggunakan beberapa metrik evaluasi untuk mengevaluasi performa model regresi linear terhadap data uji.
 - **sqrt(mean_squared_error)**: Menghitung Root Mean Squared Error (RMSE), yang mengukur seberapa baik prediksi model cocok dengan nilai sebenarnya.
 - **mean_absolute_error**: Menghitung Mean Absolute Error (MAE), yang mengukur kesalahan absolut rata-rata antara prediksi dan nilai sebenarnya.
 - **mean_squared_error**: Menghitung Mean Squared Error (MSE), yang mengukur rata-rata kuadrat kesalahan antara prediksi dan nilai sebenarnya.

Perintah `print("Linear Regression Metrics: ")`, `print(f"RMSE: {rmse_linreg} ")`, `print(f"MAE: {mae_linreg} ")`, `print(f"MSE: {mse_linreg} ")` berguna untuk menampilkan hasil evaluasi metrik untuk model regresi linear, seperti RMSE, MAE, dan MSE.

```
dataframe = pd.DataFrame({'Data Sebenarnya' : y_test, 'Data Prediksi' : y_pred_linreg})
dataframe
```

	Data Sebenarnya	Data Prediksi
DATE		
2023-03-01	33000	33062.650508
2022-02-01	31100	31208.357413
2023-01-01	38000	37960.016561
2022-10-01	33500	33484.311579
2022-04-01	30300	30301.645469
2023-06-01	36600	36626.251052
2022-08-01	39900	39855.064363

```
# Visualisasi prediksi Linear Regression
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred_linreg, color='blue')
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)],
         linestyle='--', color='red', linewidth=2)
plt.title('Linear Regression: Actual vs Predicted')
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.show()
```



0. Implementasi Model Random Forest

```
# Inisialisasi model Random Forest
rf_model = RandomForestRegressor()

# Latih model Random Forest
rf_model.fit(X_train, y_train)

# Prediksi menggunakan data uji
y_pred_rf = rf_model.predict(X_test)

# Evaluasi model Random Forest
rmse_rf = sqrt(mean_squared_error(y_test, y_pred_rf))
mae_rf = mean_absolute_error(y_test, y_pred_rf)
mse_rf = mean_squared_error(y_test, y_pred_rf)

print("Random Forest Metrics:")
print(f"RMSE: {rmse_rf}")
print(f"MAE: {mae_rf}")
print(f"MSE: {mse_rf}")
```

```
Random Forest Metrics:
RMSE: 753.3638847576534
MAE: 590.8571428571429
MSE: 567557.1428571428
```

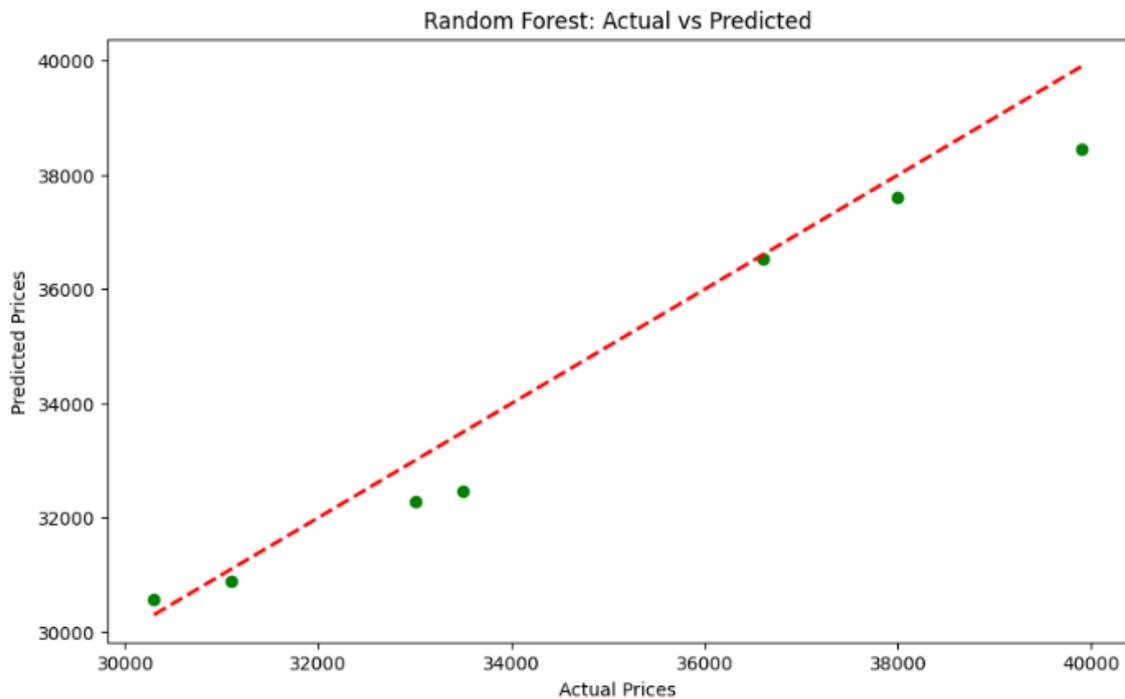
Perintah `pd.DataFrame()` digunakan untuk membuat DataFrame, Data Sebenarnya dan Data Prediksi Nama kolom pada DataFrame. Variabel `y_test` yang berisi data aktual yang biasanya digunakan sebagai variabel dependen dalam evaluasi model. Variabel `y_pred_rf` yang berisi data prediksi yang dihasilkan oleh model Random Forest.

DataFrame akan memiliki dua kolom, 'Data Sebenarnya' dan 'Data Prediksi', di mana setiap baris merepresentasikan pasangan nilai aktual dan nilai prediksi untuk suatu observasi. Dengan mencetak DataFrame ini, Anda dapat melihat perbandingan antara data aktual dan data prediksi untuk memahami sejauh mana model berhasil dalam melakukan prediksi.

```
dataframe = pd.DataFrame({'Data Sebenarnya' : y_test, 'Data Prediksi' : y_pred_rf})
dataframe
```

	Data Sebenarnya	Data Prediksi
DATE		
2023-03-01	33000	32530.0
2022-02-01	31100	30977.0
2023-01-01	38000	37603.0
2022-10-01	33500	32913.0
2022-04-01	30300	30579.0
2023-06-01	36600	36420.0
2022-08-01	39900	37784.0

```
# Visualisasi prediksi Random Forest
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred_rf, color='green')
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], linestyle='--',
         color='red', linewidth=2)
plt.title('Random Forest: Actual vs Predicted')
plt.xlabel('Actual Prices')
plt.ylabel('Predicted Prices')
plt.show()
```



Performa model Linear Regression memiliki RMSE, MAE, MSE, lebih rendah dan R-squared yang lebih tinggi dibandingkan dengan Random Forest, maka dapat disimpulkan bahwa Linear Regression memiliki performa lebih baik. Visualisasi hasil menggunakan Scatter plot membantu melihat sejauh mana nilai prediksi model mendekati nilai aktual. Jika titik-titik pada scatter plot berdekatan dengan garis diagonal, model cenderung memiliki prediksi yang baik. Berdasarkan hasil eksperimen di atas terlihat nilai RMSE pada Linear Regression lebih kecil yaitu 53.74842694081432 dari Random Forest yaitu 753.3638847576534, sehingga hasil evaluasi dan visualisasi, dapat disimpulkan model yang lebih baik untuk kasus ini adalah Metode Linear Regresi.

KESIMPULAN

Analisis performa kedua model menunjukkan bahwa dalam memprediksi variabel dependen pada dataset ini, model regresi linier memberikan hasil yang lebih baik dengan tingkat RMSE yang lebih rendah dibandingkan dengan model Random Forest. Tingkat RMSE yang lebih rendah menunjukkan bahwa model regresi linier cenderung membuat prediksi yang lebih akurat dan lebih dekat dengan nilai observasional sebenarnya.

Sifat-sifat Random Forest yang kompleks lebih kompleks daripada model regresi linier. Dalam beberapa situasi, model yang lebih sederhana mungkin menghasilkan hasil yang lebih baik. Ini terutama berlaku ketika hubungan antara variabel independen dan dependen bersifat linear. Hasil ini juga dapat dipengaruhi oleh ukuran dataset. Model yang lebih kompleks seperti Random Forest dapat bekerja lebih baik pada dataset tertentu, terutama jika ada pola non-linear yang kompleks. Hasil ini juga dapat dipengaruhi oleh penyetelan parameter yang ideal untuk masing-masing model; penyetelan parameter yang baik dapat meningkatkan kinerja model secara signifikan.

DAFTAR PUSTAKA

- Alfin, M.K. *et al.* (2022) 'Penerapan Metode Clustering Untuk Prediksi Produksi Bawang Merah (Ensemble K-Nearest Neighbors)', 3(2), pp. 30–37.
- Arfa, M.F. *et al.* (2023) 'Prediksi Harga Cryptocurrency Dengan Metode Linear Regresi', *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 1(1), pp. 8–15. Available at: <https://journal.irpi.or.id/index.php/sentimas/article/view/609%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/609/332>.
- Arinal, V. and Azhari, M. (2023) 'Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia', *Jurnal Sains dan Teknologi*, 5(1), p. pp. Available at: <https://doi.org/10.55338/saintek.v5i1.1417>.
- Fadhilah, M. and Ali, I. (2023) 'Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear di Nolbas SVNR', *INTERNAL (Information System Journal)*, 6(1), pp. 22–32. Available at: <http://jurnal.masoemuniveristy.ac.id/index.php/internal>.
- Fitri, E. (2023) 'Analisis Perbandingan Metode Regresi Linear, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah', *Journal of Applied Computer Science and Technology*, 4(1), pp. 58–64. Available at: <https://doi.org/10.52158/jacost.v4i1.491>.
- Musthofa, N. *et al.* (2023) 'Penerapan Algoritma Regresi Linear Sederhana Untuk Prediksi Pengaruh Nilai Omset Terhadap Besaran Pendapatan Bersih Usaha Mikro Kecil Dan ...', ... (*Jurnal Informatika dan ...*, 04(01), pp. 46–56. Available at: <https://ejurnalunsam.id/index.php/jicom/article/download/5935/4358>.
- Ong, A.S.H. (2009) 'Penerapan Metode Regresi Linear Sederhana untuk Prediksi Harga Beras di Kota Medan', *Journal of Science and Technology in the Tropics*, 5(1), p. 3.
- Puteri, K. and Silvanie, A. (2020) 'Machine Learning untuk Model Prediksi Harga Sembako', *Jurnal Nasional Informatika*, 1(2), pp. 82–94.
- Redo, M.R. *et al.* (2021) 'Perbandingan performa Algoritma Neural Network, Regresi Linear, dan Random Forest dalam simulasi prediksi angka kematian pasien COVID-19 di Indonesia', *Institut Informatika dan Bisnis Darmajaya*, pp. 54–62.
- Supriyanto, S., Ilhamsyah, M. and Enri, U. (2022) 'Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression Dan Random Forest', *Jurnal Ilmiah Wahana Pendidikan*, 8(7), pp. 1–8. Available at: <https://doi.org/10.5281/zenodo.6559603>.
- Suryanto, A.A. (2019) 'Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi', *Saintekbu*, 11(1), pp. 78–83. Available at: <https://doi.org/10.32764/saintekbu.v11i1.298>.
- Yusuf Nugroho, M. and Duta Bangsa Surakarta, U. (2023) 'Analisis Faktor-Faktor yang Mempengaruhi Produksi Padi di Sumatera Menggunakan Metode Regresi Linear', pp. 36–40.