

Penerapan Metode Random Forest Regression Untuk Menganalisis Manajemen Stok Produksi Semen di PT. Semen Tonasa

Rismayani¹, Sri Wahyuni¹, Sartin M. Bunga^{2*}

¹Universitas Dipa Makassar, Makassar, Indonesia.

²Sistem Informasi; Universitas Dipa Makassar, Makassar, Indonesia.

e-mail:¹rismayani@undipa.ac.id, ²sriwahyuni@undipa.ac.id, ³sartinbunga@gmail.com

Abstrak

Manajemen stok produksi semen merupakan aspek penting dalam mendukung kelancaran operasional dan distribusi pada unit packing plant. Ketidakseimbangan antara ketersediaan stok dan kebutuhan distribusi berpotensi menimbulkan kelebihan maupun kekurangan stok yang dapat memengaruhi efisiensi operasional. Penelitian ini bertujuan untuk menerapkan metode Random Forest Regression dalam memprediksi stok produksi semen serta menganalisis kondisi manajemen stok produksi pada Unit Packing Plant Bitung PT Semen Tonasa. Data yang digunakan merupakan data historis stok produksi semen selama dua tahun terakhir, yaitu tahun 2024 dan 2025. Metode penelitian meliputi pengolahan data, analisis eksploratif, pembangunan dan evaluasi model Random Forest Regression, analisis feature importance, serta prediksi stok produksi satu tahun mendatang. Evaluasi kinerja model dilakukan menggunakan metrik Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model memiliki kinerja yang sangat baik dengan nilai MAE sebesar 4,60, RMSE sebesar 8,21, dan R^2 sebesar 0,9983. Analisis feature importance menunjukkan bahwa variabel total bag merupakan faktor yang paling dominan dalam memengaruhi stok produksi semen. Hasil prediksi menunjukkan adanya variasi nilai stok produksi antarperiode yang mencerminkan kondisi manajemen stok produksi yang dinamis.

Kata kunci: Manajemen Stok Produksi, *Random Forest Regression*, Stok Produksi Semen, Packing Plant.

I. PENDAHULUAN

Industri semen merupakan salah satu sektor strategis yang berperan penting dalam mendukung pembangunan infrastruktur. Ketersediaan semen yang stabil sangat dipengaruhi oleh kemampuan perusahaan dalam mengelola stok produksi secara efektif. PT Semen Tonasa sebagai salah satu produsen semen terbesar di kawasan Indonesia Timur memiliki peran penting dalam menjaga ketersediaan pasokan produk semen untuk berbagai wilayah distribusi. Salah satu unit distribusi penting yang dimiliki perusahaan adalah Unit Packing Plant Bitung, yang berfungsi sebagai pusat pengemasan, penyimpanan, dan distribusi semen ke berbagai wilayah di Indonesia Timur [1].

Manajemen stok produksi semen menjadi aspek krusial dalam menjaga keseimbangan antara ketersediaan stok dan kebutuhan distribusi. Ketidaktepatan dalam pengelolaan stok dapat menimbulkan dua kondisi yang merugikan perusahaan, yaitu overstock dan stockout. Overstock dapat menyebabkan penumpukan stok di gudang yang berpotensi meningkatkan biaya penyimpanan serta menurunkan kualitas semen akibat proses penggumpalan. Sebaliknya, stockout dapat menyebabkan tidak terpenuhinya permintaan pelanggan dan menghambat proses distribusi produk. Oleh karena itu, diperlukan pendekatan analisis berbasis data untuk memahami

dinamika stok produksi dan mendukung pengambilan keputusan operasional secara lebih akurat [2].

Perkembangan teknologi machine learning memungkinkan perusahaan memanfaatkan data historis operasional untuk melakukan analisis dan prediksi yang lebih akurat dibandingkan metode konvensional. Machine learning mampu mempelajari pola hubungan antar variabel operasional sehingga dapat menghasilkan model analisis yang mampu menangkap pola data yang kompleks dan bersifat non-linear. Dalam konteks manajemen stok, penerapan machine learning dapat membantu perusahaan memahami faktor-faktor yang mempengaruhi perubahan stok serta meningkatkan efektivitas pengelolaan persediaan [3].

Salah satu metode machine learning yang banyak digunakan dalam analisis data industri adalah Random Forest Regression. Metode ini merupakan algoritma berbasis ensemble learning yang bekerja dengan membangun sejumlah decision tree secara paralel dan menggabungkan hasil prediksi dari setiap pohon keputusan. Random Forest memiliki keunggulan dalam menangani hubungan variabel yang kompleks, memiliki tingkat sensitivitas yang rendah terhadap outlier, serta mampu menghasilkan tingkat akurasi yang tinggi dalam proses prediksi. Selain itu, metode ini juga mampu menunjukkan variabel yang paling berpengaruh terhadap hasil prediksi melalui analisis

feature importance, sehingga memberikan informasi yang bermanfaat bagi pengambilan keputusan manajerial [4], [5]

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk menerapkan metode Random Forest Regression dalam menganalisis manajemen stok produksi semen pada Unit Packing Plant Bitung PT Semen Tonasa dengan memanfaatkan data historis operasional perusahaan. Melalui penerapan model ini diharapkan dapat diperoleh analisis yang akurat mengenai kondisi stok produksi serta identifikasi faktor-faktor dominan yang mempengaruhi perubahan stok, sehingga dapat memberikan kontribusi dalam meningkatkan efektivitas pengelolaan stok produksi semen di perusahaan.

II. METODOLOGI PENELITIAN

A. Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan dalam rentang waktu Desember 2025 hingga Februari 2026. Lokasi penelitian dilakukan pada PT Semen Tonasa, khususnya pada Unit Packing Plant Bitung yang beralamat di Jl. Samuel Languyu, Kelurahan Pateten Satu, Kecamatan Aertembaga, Kota Bitung, Sulawesi Utara. Unit ini berperan sebagai pusat pengemasan, penyimpanan, dan distribusi semen yang menjadi objek analisis dalam penelitian ini.

B. Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif dengan pendekatan analisis prediktif berbasis machine learning. Pendekatan kuantitatif dipilih karena penelitian memanfaatkan data historis stok produksi semen yang disajikan dalam bentuk numerik. Proses analisis dilakukan menggunakan metode komputasional untuk membangun model prediksi yang mampu menganalisis hubungan antar variabel operasional yang mempengaruhi perubahan stok produksi semen.

C. Sumber Data

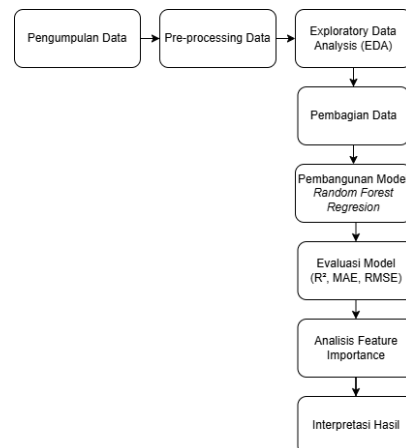
Sumber data dalam penelitian ini terdiri dari data primer dan data sekunder. Data primer diperoleh melalui komunikasi dengan pihak operasional Unit Packing Plant Bitung PT Semen Tonasa. Data ini digunakan untuk memperoleh pemahaman mengenai alur operasional, arti variabel dalam dataset, serta proses pencatatan stok produksi. Data sekunder merupakan data utama dalam penelitian yang diperoleh dalam bentuk arsip digital perusahaan. Data ini berisi catatan historis mengenai stok produksi semen, aktivitas produksi, dan variabel operasional lainnya yang berkaitan dengan pengelolaan stok produksi di Unit Packing Plant Bitung.

D. Metode Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui analisis dokumen, yaitu dengan memanfaatkan arsip data operasional perusahaan yang berkaitan dengan aktivitas stok produksi. Selain itu, penelitian juga didukung

oleh data primer yang diperoleh melalui komunikasi dengan pihak operasional untuk memahami konteks pencatatan data serta arti dari setiap variabel yang terdapat dalam dataset.

E. Prosedur Penelitian



Gambar 1. Prosedur Penelitian

Gambar 1 menunjukkan tahapan penelitian yang dilakukan dalam menganalisis manajemen stok produksi semen menggunakan metode *Random Forest Regression*. Proses penelitian dimulai dari pengumpulan data historis stok produksi semen yang diperoleh dari Unit Packing Plant Bitung PT Semen Tonasa. Tahap selanjutnya adalah *preprocessing* data, yaitu membersihkan data, mengecek *missing value*, serta menyesuaikan tipe data agar dataset siap digunakan dalam analisis. Setelah itu dilakukan eksplorasi data untuk memahami karakteristik data dan hubungan antar variabel operasional. Dataset kemudian dibagi menjadi *data training* (80%) untuk melatih model dan *data testing* (20%) untuk menguji kinerja model. Model *Random Forest Regression* dilatih menggunakan data training untuk mempelajari pola hubungan antar variabel. Selanjutnya model diuji menggunakan data testing untuk mengetahui tingkat akurasi. Tahap terakhir adalah evaluasi model menggunakan metrik MAE, RMSE, dan R^2 untuk menilai kinerja model. Selain itu dilakukan analisis *feature importance* untuk mengetahui variabel yang paling berpengaruh terhadap perubahan stok produksi semen.

III. HASIL DAN PEMBAHASAN

Pada tahap ini, data yang digunakan merupakan data historis stok produksi semen pada Unit Packing Plant Bitung PT Semen Tonasa. Data tersebut diperoleh dari catatan operasional perusahaan selama periode tahun 2024 hingga 2025. Dataset yang digunakan terdiri dari berbagai variabel yang berkaitan dengan aktivitas produksi dan distribusi semen. Sebelum digunakan dalam proses pemodelan, data terlebih dahulu melalui tahap *preprocessing* untuk memastikan kualitas data yang digunakan dalam penelitian. Tahap ini meliputi proses pembersihan data, pengecekan nilai yang hilang, serta

pemilihan variabel yang relevan dengan penelitian. Setelah proses tersebut dilakukan, diperoleh dataset yang siap digunakan dalam analisis menggunakan metode *Random Forest Regression*. Data yang digunakan dalam penelitian ini terdiri dari beberapa variabel utama yaitu tanggal, total bulk, total bag, dan total stok produksi semen. Sebagian dataset yang digunakan dalam penelitian ini ditampilkan pada Tabel 1 berikut.

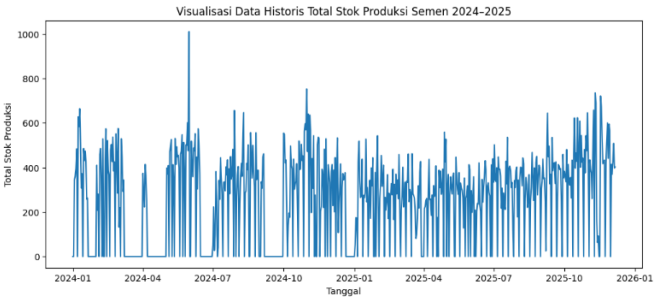
Table 1. Dataset Produksi Semen

Tanggal	Total Bulk	Total Bag	Total Stok
01/01/2024	50	295	345
02/01/2024	40	280	320
03/01/2024	45	260	305
04/01/2024	30	270	300
05/01/2024	35	250	285
06/01/2024	60	310	370
07/01/2024	55	300	355

A. Analisis Data

Pada tahap ini dilakukan analisis data awal untuk memahami karakteristik dataset yang digunakan dalam penelitian. Analisis ini bertujuan untuk melihat pola data serta hubungan antar variabel yang digunakan dalam proses pemodelan. Dataset yang telah melalui tahap preprocessing kemudian dianalisis menggunakan teknik *Exploratory Data Analysis* (EDA).

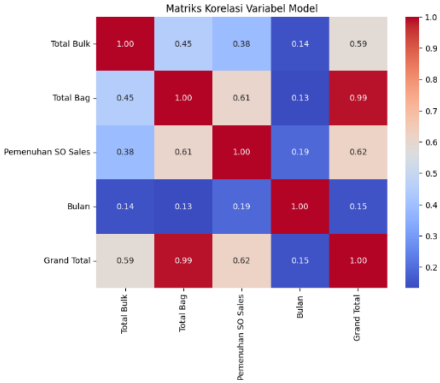
Analisis ini dilakukan dengan memvisualisasikan data stok produksi semen untuk melihat pola distribusi dan fluktuasi nilai stok selama periode pengamatan. Visualisasi data ini membantu dalam memahami karakteristik data sebelum dilakukan proses pemodelan menggunakan metode *Random Forest Regression*.



Gambar 2. Visualisasi Stok Produksi

Grafik pada Gambar 2 menunjukkan bahwa stok produksi semen mengalami fluktuasi pada setiap periode waktu. Perubahan nilai stok ini dipengaruhi oleh aktivitas produksi dan distribusi semen yang terjadi pada Unit Packing Plant Bitung PT Semen Tonasa.

Selanjutnya dilakukan analisis hubungan antar variabel yang digunakan dalam penelitian. Analisis ini dilakukan dengan menggunakan korelasi matriks untuk mengetahui tingkat hubungan antar variabel yang digunakan dalam pemodelan.



Gambar 3. Korelasi Matrix

Hasil analisis korelasi menunjukkan bahwa beberapa variabel memiliki hubungan yang cukup kuat terhadap perubahan stok produksi semen. Variabel yang memiliki hubungan kuat ini kemudian digunakan sebagai fitur dalam pembangunan model *Random Forest Regression*.

Setelah tahap analisis data dilakukan, langkah selanjutnya adalah membangun model prediksi menggunakan metode *Random Forest Regression*.

B. Penerapan Metode Random Forest Regression

Pada tahap ini dilakukan proses pembangunan model prediksi menggunakan metode Random Forest Regression. Metode ini merupakan algoritma machine learning berbasis ensemble learning yang bekerja dengan membangun sejumlah decision tree dan menggabungkan hasil prediksi dari setiap pohon keputusan untuk memperoleh hasil prediksi yang lebih stabil.

Secara matematis, model *Random Forest Regression* dapat dinyatakan sebagai berikut:

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

Dimana:

- $\hat{y}$: Nilai prediksi stok produksi semen
- T : Jumlah pohon keputusan dalam model *Random Forest*
- $f_t(x)$: hasil prediksi dari pohon keputusan ke - t
- (x) : Variabel input yang digunakan dalam model

Dalam penelitian ini, dataset yang telah melalui proses preprocessing kemudian dibagi menjadi dua bagian yaitu data training dan data testing. Data *training* digunakan untuk melatih model agar dapat mempelajari pola hubungan antar variabel, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya.

1) Perhitungan Random Forest Regression

Untuk menjelaskan proses perhitungan *Random Forest Regression* berikut perhitungan menggunakan satu sampel data. Pada salah satu decision tree, data uji masuk ke leaf node yang berisi beberapa nilai target stok produksi sebagai berikut:

Table 2. Leaf Node

Data pada Leaf Node
220
240
235
232
232

a. Langkah pertama adalah menjumlahkan seluruh nilai target pada leaf node

$$220 + 240 + 235 + 232 + 232 = 1159$$

b. Langkah kedua menghitung rata-rata leaf node

$$\frac{1159}{5} = 231.8$$

c. Nilai prediksi tree

$$\hat{y} = 231.8$$

Berdasarkan hasil perhitungan di atas, diperoleh nilai prediksi sebesar 231.8 dari salah satu decision tree. Nilai prediksi tersebut diperoleh dari rata-rata nilai target yang terdapat pada leaf node. Setiap decision tree dalam metode *Random Forest Regression* menghasilkan nilai prediksi yang berbeda sesuai dengan data yang masuk pada leaf node. Selanjutnya seluruh hasil prediksi dari setiap decision tree digabungkan dengan menggunakan proses perataan (*averaging*) untuk menghasilkan nilai prediksi akhir dari model Random Forest.

C. Evaluasi Model

Table 3. Hasil Evaluasi Model

Metrik Evaluasi	Nilai
MAE	4.60
RMSE	8.21
R ²	0.9983

Berdasarkan hasil evaluasi model pada Tabel 2, dapat diketahui bahwa model *Random Forest Regression* memiliki kinerja yang sangat baik dalam memprediksi stok produksi semen. Nilai MAE dan RMSE menunjukkan tingkat kesalahan prediksi yang relatif kecil, sedangkan nilai koefisien determinasi (R²) sebesar 0.9983 menunjukkan bahwa model mampu menjelaskan hampir seluruh variasi data stok produksi pada dataset penelitian.

1) Evaluasi Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

y_i : nilai actual

$\hat{y}_i$: nilai hasil prediksi

n : jumlah data uji

Sampel Perhitungan Manual :

$$|y_1 - \hat{y}_1| = |385.351 - 381.90| = 3.451$$

$$|y_2 - \hat{y}_2| = |0 - 4.20| = 4.200$$

$$|y_3 - \hat{y}_3| = |0 - 3.10| = 3.100$$

$$|y_4 - \hat{y}_4| = |493.7121 - 489.6000| = 4.1121$$

$$|y_5 - \hat{y}_5| = |541.8291 - 536.8000| = 5.0291$$

$$|y_6 - \hat{y}_6| = |177.2556 - 182.0000| = 4.7444$$

$$|y_7 - \hat{y}_7| = |462.03215 - 457.50000| = 4.53215$$

$$|y_8 - \hat{y}_8| = |393.55315 - 388.90000| = 4.65315$$

$$|y_9 - \hat{y}_9| = |320.6894 - 316.2000| = 4.4894$$

$$|y_{10} - \hat{y}_{10}| = |0 - 4.000| = 4.000$$

$$MAE_{(10 \text{ data})} = \frac{42.31}{10} = 4.231$$

$$MAE = \frac{1}{142} \sum_{i=1}^{142} |y_i - \hat{y}_i| = 4.900832394366166$$

dimana y_i merupakan nilai actual, $\hat{y}_i$ merupakan nilai hasil prediksi, dan n adalah jumlah data uji. Untuk menunjukkan proses perhitungan MAE berdasarkan rumus tersebut, dilakukan perhitungan manual menggunakan 10 data uji sebagai sampel. Pada setiap data uji dihitung selisih absolut antara nilai aktual dan nilai prediksi. Seluruh selisih absolut dari 10 data tersebut kemudian dijumlahkan sehingga diperoleh total selisih absolut sebesar 42,31. Nilai ini selanjutnya dibagi dengan jumlah data contoh, yaitu 10 data, sehingga diperoleh nilai MAE contoh sebesar 4,231. Perhitungan manual tersebut digunakan untuk menunjukkan mekanisme perhitungan MAE secara matematis. Adapun nilai MAE akhir pada penelitian ini dihitung menggunakan seluruh data uji sebanyak 142 data dengan bantuan *Google Colab*. Berdasarkan hasil perhitungan tersebut diperoleh nilai MAE sebesar 4,900832394366166. Nilai ini menunjukkan bahwa secara rata-rata, hasil prediksi model memiliki selisih sekitar 4,90 terhadap nilai aktual total stok produksi semen.

2) Evaluasi Mean Squared Error (MSE)

Mean Squared Error (MSE) digunakan untuk mengukur rata-rata kuadrat kesalahan antara nilai aktual dan nilai hasil prediksi yang dihasilkan oleh model *Random Forest Regression*. MSE menghitung selisih antara nilai aktual dan nilai prediksi pada setiap data uji, kemudian mengkuadratkan selisih tersebut. Pengkuadratan dilakukan untuk memberikan penalti yang lebih besar terhadap kesalahan prediksi dengan selisih yang besar.

Secara matematis, Mean Squared Error (MSE) dirumuskan sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Dimana:

y_i : nilai actual
 $\hat{y}_i$: nilai hasil prediksi
 n : jumlah data uji

Untuk menunjukkan proses perhitungan MSE berdasarkan rumus tersebut, dilakukan perhitungan manual menggunakan 10 data uji sebagai contoh. Pada setiap data uji dihitung kuadrat selisih antara nilai aktual dan nilai prediksi. Hasil perhitungan kuadrat selisih untuk masing-masing data uji kemudian dijumlahkan.

Berdasarkan perhitungan manual menggunakan 10 data uji, diperoleh jumlah kuadrat selisih sebesar 182,207. Nilai tersebut kemudian dibagi dengan jumlah data contoh, yaitu 10 data, sehingga diperoleh nilai MSE contoh sebagai berikut:

1. Sampel Perhitungan Manual

$$\begin{aligned}(y_1 - \hat{y}_1)^2 &= (385.351 - 381.900)^2 = 11.909 \\(y_2 - \hat{y}_2)^2 &= (0 - 4.200)^2 = 17.640 \\(y_3 - \hat{y}_3)^2 &= (0 - 3.100)^2 = 9.610 \\(y_4 - \hat{y}_4)^2 &= (493.7121 - 489.6000)^2 = 16.910 \\(y_5 - \hat{y}_5)^2 &= (541.8291 - 536.8000)^2 = 25.292 \\(y_6 - \hat{y}_6)^2 &= (177.2556 - 182.0000)^2 = 22.506 \\(y_7 - \hat{y}_7)^2 &= (462.03215 - 457.50000)^2 = 20.539 \\(y_8 - \hat{y}_8)^2 &= (393.55315 - 388.90000)^2 = 21.646 \\(y_9 - \hat{y}_9)^2 &= (320.6894 - 316.2000)^2 = 20.155 \\(y_{10} - \hat{y}_{10})^2 &= (0 - 4.000)^2 = 16.000\end{aligned}$$

2. Penjumlahan Kuadrat Error:

$$\begin{aligned}\sum_{i=1}^{10} (y_i - \hat{y}_i)^2 &= 11.909 + 17.640 + 9.610 + 16.910 \\&\quad + 25.292 + 22.506 + 20.539 \\&\quad + 21.646 + 20.155 + 16.000 \\&= 182.207\end{aligned}$$

3. Hasil MSE dari sampel diatas :

$$MSE_{(10 \text{ data})} = \frac{182.207}{10} = 18.2207$$

4. Hasil MSE Final Keseluruhan Data Uji – sesuai colab

$$5. \text{ MSE} = (8.892570610546485)^2 = 79.08$$

Perhitungan MSE dilakukan dengan mengkuadratkan selisih antara nilai aktual dan nilai prediksi pada setiap data uji, kemudian menjumlahkannya dan membaginya dengan jumlah data uji. Untuk menunjukkan proses perhitungan MSE secara manual, digunakan contoh sebanyak 10 data uji. Pada setiap data uji dihitung kuadrat selisih antara nilai aktual dan nilai prediksi. Seluruh hasil kuadrat selisih tersebut dijumlahkan sehingga diperoleh total sebesar 182,207, kemudian dibagi dengan jumlah data contoh, yaitu 10 data, sehingga diperoleh nilai MSE contoh sebesar 18,2207.

Berdasarkan hasil evaluasi tersebut, diperoleh nilai MSE yang konsisten dengan hasil evaluasi model, yang selanjutnya digunakan sebagai dasar perhitungan nilai *Root Mean Squared Error* (RMSE).

3) Evaluasi *Root Mean Squared Error* (RMSE)

Root Mean Squared Error (RMSE) digunakan untuk mengukur besarnya kesalahan prediksi model dengan satuan yang sama dengan data asli. RMSE diperoleh dari akar kuadrat nilai *Mean Squared Error* (MSE), sehingga metrik ini lebih mudah diinterpretasikan karena berada pada satuan yang sama dengan variabel target.

Secara matematis, RMSE dirumuskan sebagai berikut:

$$RMSE = \sqrt{MSE}$$

1. Selisih Nilai Aktual dan Prediksi

$$\begin{aligned}y_1 - \hat{y}_1 &= 385.351 - 381.900 = 3.451 \\y_2 - \hat{y}_2 &= 0 - 4.200 = -4.200 \\y_3 - \hat{y}_3 &= 0 - 3.100 = -3.100 \\y_4 - \hat{y}_4 &= 493.7121 - 489.6000 = 4.1121 \\y_5 - \hat{y}_5 &= 541.8291 - 536.8000 = 5.0291 \\y_6 - \hat{y}_6 &= 177.2556 - 182.0000 = -4.7444 \\y_7 - \hat{y}_7 &= 462.03215 - 457.50000 = 4.53215 \\y_8 - \hat{y}_8 &= 393.55315 - 388.90000 = 4.65315 \\y_9 - \hat{y}_9 &= 320.6894 - 316.2000 = 4.4894 \\y_{10} - \hat{y}_{10} &= 0 - 4.000 = -4.000\end{aligned}$$

2. Kuadrat Selisih

$$\begin{aligned}(y_1 - \hat{y}_1)^2 &= (3.451)^2 = 11.909 \\(y_2 - \hat{y}_2)^2 &= (-4.200)^2 = 17.640 \\(y_3 - \hat{y}_3)^2 &= (-3.100)^2 = 9.610 \\(y_4 - \hat{y}_4)^2 &= (4.1121)^2 = 16.910 \\(y_5 - \hat{y}_5)^2 &= (5.0291)^2 = 25.292 \\(y_6 - \hat{y}_6)^2 &= (-4.7444)^2 = 22.506 \\(y_7 - \hat{y}_7)^2 &= (4.53215)^2 = 20.539 \\(y_8 - \hat{y}_8)^2 &= (4.65315)^2 = 21.646 \\(y_9 - \hat{y}_9)^2 &= (4.4894)^2 = 20.155 \\(y_{10} - \hat{y}_{10})^2 &= (-4.000)^2 = 16.000\end{aligned}$$

3. Penjumlahan Kuadrat Selisih

$$\begin{aligned}\sum_{i=1}^{10} (y_i - \hat{y}_i)^2 &= 11.909 + 17.640 + 9.610 + 16.910 \\&\quad + 25.292 + 22.506 + 20.539 \\&\quad + 21.646 + 20.155 + 16.000 \\&= 182.207\end{aligned}$$

4. Hitung MSE (10 Data)

$$MSE_{(10 \text{ data})} = \frac{182.207}{10} = 18.2207$$

5. Hitung RMSE (10 Data)

$$RMSE_{(10 \text{ data})} = \sqrt{18.2207} = 4.268$$

6. RMSE Akhir (Seluruh Data Uji – sesuai Colab)

$$RMSE = \sqrt{79.08} = 8.892570610546485$$

Proses Perhitungan Manual RMSE:

Berdasarkan hasil perhitungan *Mean Squared Error* (MSE) menggunakan 10 data uji sebagai contoh pada tahap sebelumnya, diperoleh nilai MSE contoh sebesar 18,2207. Nilai

RMSE contoh kemudian dihitung dengan mengambil akar kuadrat dari nilai MSE tersebut, sebagai berikut:

$$MSE_{(10 \text{ data})} = \frac{182.207}{10} = 18.2207$$

Perhitungan ini menunjukkan mekanisme perolehan nilai RMSE berdasarkan rumus yang digunakan. Nilai RMSE contoh tersebut hanya digunakan untuk memperlihatkan proses perhitungan secara manual dan tidak merepresentasikan nilai RMSE akhir penelitian.

Nilai RMSE Akhir (Seluruh Data Uji)

Nilai RMSE akhir pada penelitian ini dihitung menggunakan seluruh data uji sebanyak 142 data dengan bantuan *Google Colab*. Berdasarkan hasil evaluasi model, diperoleh nilai RMSE sebagai berikut:

$$RMSE = \sqrt{79.08} = 8.892570610546485$$

Nilai RMSE tersebut menunjukkan bahwa secara rata-rata, kesalahan prediksi model terhadap nilai aktual total stok produksi semen berada pada kisaran 8,89 satuan. Nilai ini tergolong rendah dan mengindikasikan bahwa model *Random Forest Regression* memiliki kemampuan prediksi yang baik dan stabil terhadap data uji.

4) Evaluasi Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai R^2 menunjukkan seberapa besar proporsi variasi pada variabel target yang dapat dijelaskan oleh model prediksi. Nilai R^2 berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang semakin baik.

Rumus Koefisien Determinasi (R^2) :

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

Dimana :

y_i : nilai actual

$\hat{y}_i$: nilai hasil prediksi

$\bar{y}$: rata-rata nilai actual

n : jumlah data uji

Proses Perhitungan R^2 :

Perhitungan R^2 dilakukan dengan membandingkan dua komponen utama, yaitu:

1. Jumlah Kuadrat Kesalahan Prediksi (SSE)

Jumlah kuadrat kesalahan prediksi (*Sum of Squared Errors* / SSE) dihitung dengan menjumlahkan kuadrat selisih antara nilai aktual dan nilai prediksi pada seluruh data uji, yaitu:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Pada penelitian ini, nilai SSE diperoleh dari hasil evaluasi model menggunakan *Google Colab*. Nilai SSE tersebut merupakan total kesalahan prediksi model terhadap seluruh data uji.

2. Jumlah Kuadrat variasi data Aktual (SST)

Jumlah kuadrat variasi data aktual (*Total Sum of Squares* / SST) dihitung dengan menjumlahkan kuadrat selisih antara nilai aktual dan nilai rata-ratanya, yaitu:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

Nilai SST merepresentasikan total variasi data aktual terhadap nilai rata-ratanya.

3. Substitusi kedalam rumus R^2

Setelah nilai SSE dan SST diperoleh, nilai koefisien determinasi dihitung dengan mensubstitusikan kedua nilai tersebut ke dalam rumus R^2 :

$$R^2 = 1 - \frac{SSE}{SST}$$

Berdasarkan hasil perhitungan menggunakan *Google Colab*, diperoleh nilai koefisien determinasi sebagai berikut:

$$R^2 = 0.9980321210072012$$

4. Interpretasi Hasil R^2

Nilai R^2 sebesar 0,9980 menunjukkan bahwa sekitar 99,80% variasi data total stok produksi semen dapat dijelaskan oleh model *Random Forest Regression*, sedangkan sisanya sekitar 0,20% dipengaruhi oleh faktor lain di luar model. Nilai R^2 yang sangat mendekati 1 mengindikasikan bahwa model memiliki tingkat kecocokan yang sangat tinggi terhadap data aktual.

Berdasarkan hasil evaluasi model *Random Forest Regression* yang telah dilakukan, dapat disimpulkan bahwa model memiliki kinerja yang sangat baik dalam menganalisis manajemen stok produksi semen. Nilai *Mean Absolute Error* (MAE) sebesar 4,9008 menunjukkan bahwa secara rata-rata selisih absolut antara nilai prediksi dan nilai aktual relatif kecil. Nilai *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE) yang diperoleh, yaitu MSE $\approx$ 79,08 dan RMSE sebesar 8,8926, menunjukkan bahwa kesalahan prediksi model masih berada dalam batas yang dapat diterima dan stabil terhadap variasi data.

Selain itu, nilai koefisien determinasi (R^2) sebesar 0,9980 menunjukkan bahwa model mampu menjelaskan sekitar 99,80% variasi data total stok produksi semen. Nilai R^2 yang mendekati 1 mengindikasikan bahwa hubungan antara variabel input dan variabel target dapat direpresentasikan dengan sangat baik oleh model yang dibangun.

Dengan demikian, berdasarkan hasil evaluasi menggunakan metrik MAE, MSE, RMSE, dan R^2 , model *Random Forest Regression* yang dibangun dinilai akurat, stabil, dan layak digunakan untuk menganalisis manajemen stok produksi semen pada unit packing plant. Hasil evaluasi ini menunjukkan bahwa model telah memenuhi tujuan penelitian dan dapat dijadikan dasar untuk pembahasan dan penarikan kesimpulan pada bab selanjutnya.

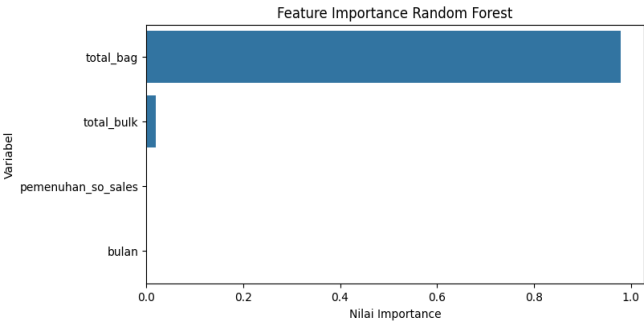
5. Analisis Feature Importance

Berdasarkan hasil pelatihan model *Random Forest Regression* menggunakan *Google Colab*, diperoleh nilai

feature importance untuk masing-masing variabel input sebagaimana ditunjukkan pada Tabel dan Gambar berikut:

Table 4. Feature Importance

Fitur	Nilai <i>Importance</i>
Total Bag	0.969814
Total Bulk	0.019310
Pemenuhan SO Sales	0.010277
Bulan	0.000600



Gambar 4. Feature Importance

Berdasarkan tabel tersebut, terlihat bahwa variabel Total Bag memiliki nilai *feature importance* yang sangat dominan, yaitu sebesar 0,969814, atau mendekati 1. Sementara itu, variabel Total Bulk, Pemenuhan SO Sales, dan Bulan memiliki nilai *feature importance* yang jauh lebih kecil dibandingkan Total Bag.

Hasil ini juga diperkuat oleh visualisasi grafik *feature importance*, di mana batang grafik untuk variabel Total Bag terlihat jauh lebih panjang dibandingkan variabel lainnya. Hal ini menunjukkan bahwa kontribusi variabel Total Bag terhadap prediksi total stok produksi semen jauh lebih besar dibandingkan kontribusi variabel input lainnya.

6. *Prediksi Stok Produksi Semen 1 Tahun Mendatang*

Setelah model *Random Forest Regression* dibangun dan dievaluasi, tahap selanjutnya adalah melakukan prediksi terhadap stok produksi semen untuk periode satu tahun ke depan. Prediksi ini dilakukan dengan memanfaatkan model yang telah dilatih menggunakan data historis stok produksi semen pada Unit Packing Plant Bitung PT Semen Tonasa.

Model *Random Forest Regression* yang telah dilatih kemudian digunakan untuk menghasilkan nilai prediksi stok produksi semen pada periode selanjutnya berdasarkan pola hubungan antar variabel yang telah dipelajari dari data historis. Hasil prediksi ini memberikan gambaran mengenai kemungkinan perubahan stok produksi semen pada periode mendatang.

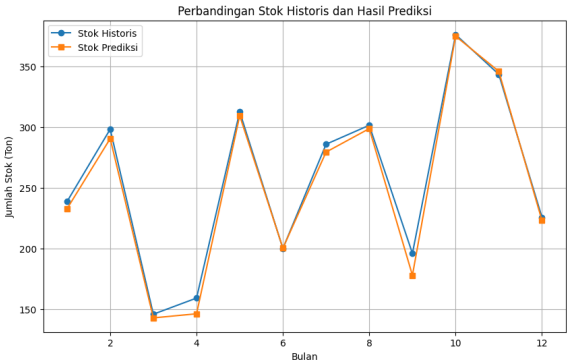
Prediksi stok produksi satu tahun mendatang diperoleh melalui pengelompokan data historis berdasarkan variabel bulan untuk menghitung nilai rata-rata total stok produksi (Grand Total) pada masing-masing bulan. Hasil pengelompokan tersebut menghasilkan tabel yang menunjukkan nilai rata-rata stok produksi untuk bulan 1 hingga

bulan 12, di mana setiap bulan memiliki nilai Grand Total yang berbeda. Perbedaan nilai ini menunjukkan bahwa berdasarkan data historis yang digunakan, tingkat stok produksi tidak berada pada nilai yang sama di setiap bulan, melainkan bervariasi sesuai dengan karakteristik data yang tercatat sebelumnya. Nilai rata-rata stok produksi bulanan tersebut kemudian digunakan sebagai dasar pembentukan tabel prediksi stok produksi satu tahun mendatang dengan menyesuaikan penamaan kolom menjadi Prediksi Stok Produksi.

Table 5. Prediksi Stok

Bulan	Stok Historis	Stok Prediksi
1.0	238.530161	232.58280
2.0	298.166842	290.42135
3.0	145.812258	142.76010
4.0	159.084000	146.15200
5.0	312.613226	309.26305
6.0	199.908333	200.21785
7.0	285.801290	279.32370
8.0	301.409839	298.65610
9.0	195.882833	177.85705
10.0	376.128548	374.80480
11.0	343.410833	345.90085
12.0	225.657568	222.97490

Berdasarkan tabel perbandingan tersebut, terlihat bahwa nilai stok prediksi memiliki kedekatan dengan nilai stok historis pada setiap bulan. Sebagai contoh, pada bulan ke-1 stok historis sebesar 238,53 ton, sedangkan hasil prediksi sebesar 232,58 ton. Pada bulan ke-10 stok historis sebesar 376,13 ton dan model menghasilkan prediksi sebesar 374,80 ton. Perbedaan yang relatif kecil antara kedua nilai tersebut menunjukkan bahwa model mampu melakukan estimasi dengan tingkat deviasi yang rendah.



Gambar 5. Visualisasi Prediksi

Visualisasi grafik perbandingan memperlihatkan bahwa garis stok prediksi mengikuti pergerakan nilai stok historis secara konsisten sepanjang dua belas bulan. Tidak terdapat

penyimpangan ekstrem antara nilai aktual dan nilai prediksi, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap karakteristik data.

Kedekatan antara stok historis dan stok prediksi ini sejalan dengan hasil evaluasi model sebelumnya, di mana diperoleh nilai: MAE sebesar 4,90, RMSE sebesar 8,89, R^2 sebesar 0,998. Nilai R^2 yang mendekati 1 menunjukkan bahwa variasi stok produksi dapat dijelaskan dengan sangat baik oleh variabel input dalam model. Sementara itu, nilai MAE dan RMSE yang relatif kecil mengindikasikan bahwa rata-rata kesalahan prediksi berada dalam batas yang rendah dibandingkan dengan skala nilai stok produksi yang berkisar antara 100–400 ton.

Dengan demikian, hasil prediksi satu tahun ke depan dapat digunakan sebagai dasar analisis manajemen stok produksi pada Unit Packing Plant Bitung. Informasi ini dapat membantu dalam perencanaan distribusi, pengendalian persediaan, serta pengambilan keputusan operasional yang lebih terukur.

Dengan demikian, kesesuaian antara tabel, grafik, dan metrik evaluasi menunjukkan bahwa model *Random Forest Regression* yang dibangun mampu memberikan estimasi stok produksi yang akurat dan stabil untuk periode satu tahun ke depan.

7. Interpretasi Manajemen Stok Produksi

Berdasarkan hasil evaluasi model, *Random Forest Regression* menunjukkan kinerja yang sangat baik dalam memprediksi total stok produksi semen. Hal ini dibuktikan dengan nilai *Mean Absolute Error* (MAE) sebesar 4,9008 dan *Root Mean Squared Error* (RMSE) sebesar 8,8926, yang menunjukkan bahwa selisih antara nilai prediksi dan nilai aktual relatif kecil. Selain itu, nilai koefisien determinasi (R^2) sebesar 0,9980 menunjukkan bahwa model mampu menjelaskan hampir seluruh variasi data stok produksi. Hasil ini menjawab pertanyaan penelitian mengenai kemampuan model *Random Forest Regression* dalam menganalisis stok produksi, serta menunjukkan bahwa metode ini layak digunakan sebagai alat analisis dalam manajemen stok produksi semen.

Jika dibandingkan dengan penelitian-penelitian sebelumnya yang menggunakan metode pembelajaran mesin dalam analisis produksi dan stok, hasil penelitian ini menunjukkan kesesuaian bahwa *Random Forest Regression* memiliki keunggulan dalam menangani hubungan nonlinier dan data dengan banyak variabel input. Beberapa penelitian terdahulu menyebutkan bahwa *Random Forest Regression* mampu memberikan akurasi yang tinggi karena menggabungkan hasil prediksi dari banyak pohon keputusan, sehingga lebih stabil dibandingkan metode regresi tunggal. Temuan pada penelitian ini menguatkan pandangan tersebut, karena model yang dibangun mampu menghasilkan tingkat kesalahan yang rendah dan tingkat kecocokan yang sangat tinggi terhadap data aktual.

Selanjutnya, hasil analisis *feature importance* memberikan temuan penting dalam konteks manajemen stok produksi. Variabel Total Bag terbukti memiliki nilai *feature importance* yang sangat dominan dibandingkan variabel lainnya. Temuan

ini memberikan bukti bahwa pengelolaan stok produksi semen di unit packing plant Bitung lebih banyak dipengaruhi oleh jumlah produksi dalam bentuk bag dibandingkan faktor lainnya. Hasil ini sejalan dengan karakteristik operasional packing plant, di mana sebagian besar aktivitas produksi dan distribusi difokuskan pada produk semen dalam kemasan. Dengan demikian, hasil penelitian ini memberikan temuan baru yang bermanfaat bagi manajemen, yaitu bahwa pengendalian Total Bag merupakan kunci utama dalam menjaga kestabilan stok produksi semen.

Di sisi lain, hasil *Exploratory Data Analysis* (EDA) sebelumnya menunjukkan bahwa variabel Total Bulk memiliki hubungan dengan total stok produksi. Namun, ketika seluruh variabel dianalisis secara simultan menggunakan *Random Forest Regression*, kontribusi Total Bulk terbukti lebih kecil dibandingkan Total Bag. Hal ini menunjukkan bahwa meskipun Total Bulk relevan secara hubungan dua variabel, variabel tersebut tidak bersifat dominan dalam konteks pemodelan multivariat. Temuan ini mengungkapkan perbedaan antara hasil analisis eksploratif dan hasil pemodelan, yang menjadi wawasan penting dalam memahami bagaimana faktor-faktor produksi berkontribusi terhadap stok secara keseluruhan.

Selain itu, hasil prediksi stok produksi satu tahun mendatang memberikan gambaran estimatif mengenai kondisi stok produksi pada masing-masing bulan. Variasi nilai prediksi antarbulan menunjukkan bahwa stok produksi bersifat dinamis dan memerlukan pengelolaan yang adaptif. Hasil ini dapat dimanfaatkan oleh pihak manajemen untuk melakukan perencanaan produksi, pengaturan distribusi, serta pengelolaan kapasitas penyimpanan secara lebih terarah. Temuan ini memperluas pemanfaatan hasil model tidak hanya sebagai alat prediksi, tetapi juga sebagai dasar analisis dalam pengambilan keputusan manajerial.

Dalam proses penelitian ini juga ditemukan beberapa keterbatasan dan masalah baru. Salah satu temuan yang muncul adalah bahwa kontribusi variabel Bulan terhadap prediksi stok produksi sangat kecil. Hal ini menunjukkan bahwa informasi bulan saja tidak cukup kuat untuk menjelaskan variasi stok produksi tanpa didukung oleh variabel operasional lainnya. Temuan ini membuka peluang bagi penelitian selanjutnya untuk menambahkan variabel lain yang lebih merepresentasikan kondisi operasional, seperti kapasitas produksi atau faktor permintaan pasar, guna meningkatkan kedalaman analisis manajemen stok.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, metode *Random Forest Regression* dapat diterapkan untuk menganalisis manajemen stok produksi semen pada Unit Packing Plant Bitung PT Semen Tonasa dengan memanfaatkan data historis operasional tahun 2024 hingga 2025. Model yang dibangun mampu mempelajari pola hubungan antara variabel operasional dengan stok produksi semen sehingga dapat menghasilkan prediksi yang akurat.

Hasil evaluasi model menunjukkan bahwa *Random Forest Regression* memiliki kinerja yang sangat baik dengan nilai

Mean Absolute Error (MAE) sebesar 4.60, Root Mean Squared Error (RMSE) sebesar 8.21, dan koefisien determinasi (R^2) sebesar 0.9983. Nilai R^2 yang mendekati angka 1 menunjukkan bahwa model mampu menjelaskan hampir seluruh variasi data stok produksi semen yang terdapat pada dataset penelitian.

Selain itu, hasil analisis feature importance menunjukkan bahwa variabel total bag merupakan variabel yang paling dominan dalam mempengaruhi perubahan stok produksi semen dibandingkan variabel lainnya. Hal ini menunjukkan bahwa aktivitas pengemasan semen dalam bentuk bag memiliki peran penting dalam dinamika stok produksi pada Unit Packing Plant Bitung PT Semen Tonasa.

Dengan demikian, penerapan metode Random Forest Regression dapat digunakan sebagai pendekatan analisis berbasis data untuk memahami pola perubahan stok produksi semen serta memberikan informasi yang dapat mendukung pengambilan keputusan dalam pengelolaan stok produksi.

V. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan metode Random Forest Regression terbukti mampu memberikan hasil analisis yang sangat baik dalam memprediksi stok produksi semen pada Unit Packing Plant Bitung PT Semen Tonasa. Untuk penelitian selanjutnya, disarankan agar pengembangan model dapat dilakukan dengan menggunakan dataset yang lebih luas serta menambahkan variabel lain yang berkaitan dengan aktivitas produksi dan distribusi semen sehingga model prediksi dapat memberikan hasil yang lebih komprehensif.

Selain itu, penelitian selanjutnya juga dapat melakukan perbandingan dengan metode machine learning lainnya untuk mengetahui metode yang paling optimal dalam menganalisis manajemen stok produksi semen. Dengan pengembangan tersebut diharapkan hasil analisis yang diperoleh dapat memberikan kontribusi yang lebih besar dalam mendukung proses pengambilan keputusan terkait pengelolaan stok produksi pada perusahaan.

VI. SARAN

Penulis mengucapkan terima kasih kepada Universitas Dipa Makassar yang telah memberikan kesempatan dan dukungan dalam pelaksanaan penelitian ini. Penulis juga menyampaikan terima kasih kepada para dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan selama proses penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada orang tua dan keluarga yang selalu memberikan doa, dukungan, dan semangat selama proses penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu baik secara langsung maupun tidak langsung sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- [1] F. Faturrahmi, "Dampak Aktivitas PT. Semen Tonasa Terhadap Lingkungan Desa Biringere Kabupaten Pangkajene Dan Kepulauan," *J. Environ. Sci.*, vol. 1, no. 1, 2018, doi: 10.35580/jes.v1i1.7342.
- [2] S. J. Fayaz, N. Montiel-Bohórquez, S. Bishnoi, M. Romano, M. Gatti, and N. M. A. Krishnan, "Industrial-scale prediction of cement clinker phases using machine learning," *Commun. Eng.*, vol. 4, no. 1, 2025, doi: 10.1038/s44172-025-00432-3.
- [3] N. Samsuddin, M. Mohd Nor, and M. Sulaiman, "Predictive Inventory Analytics and Demand Forecasting in Malaysian Electronics Manufacturing," *J. Technol. Manag. Technopreneursh.*, vol. 12, no. 2, pp. 15–24, 2024, doi: 10.54554/jtmt.2024.12.02.002.
- [4] H. Zermane, H. Madjour, A. Ziar, and A. Zermane, "Forecasting material quantity using machine learning and times series techniques," *J. Electr. Eng.*, vol. 75, no. 3, pp. 237–248, 2024, doi: 10.2478/jee-2024-0029.
- [5] P. Wang, "Research on Product Sales Forecasting Model Based on Machine Learning," *Front. Artif. Intell. Appl.*, vol. 373, pp. 1130–1135, 2023, doi: 10.3233/FAIA230929.
- [6] S. Audina and A. Bakhtiar, "Analisis Pengendalian Persediaan Aux Raw Material Menggunakan Metode Min-Max Stock Di Pt. Mitsubishi Chemical Indonesia," *J@ti Undip J. Tek. Ind.*, vol. 16, no. 3, pp. 161–168, 2021, doi: 10.14710/jati.16.3.161-168.
- [7] R. Hidayat *et al.*, "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Simkom*, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.
- [8] V. Plevris, G. Solorzano, N. P. Bakas, and M. E. A. Ben Seghier, "Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models," *World Congr. Comput. Mech. ECCOMAS Congr.*, no. June, p. 2022, 2022, doi: 10.23967/eccomas.2022.155.
- [9] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [10] R. Y. H. Silitonga, L. R. Kristiana, and T. A. Parley, "A Multi-Item Probabilistic Inventory Model that Considers Expiration Factor, All Unit Discount Policy and Warehouse Capacity Constraints," *J. Tek. Ind.*, vol. 23, no. 2, pp. 139–148, 2021, doi: 10.9744/jti.23.2.139-148.
- [11] M. Y. Urrochman, H. Asy'ari, and F. A. Hizham, "Performance Comparison of Random Forest Regression, Svr Models in Stock Price Prediction," *J. Pilar Nusa Mandiri*, vol. 21, no. 1, pp. 16–23, 2025, doi: 10.33480/pilar.v21i1.6072.
- [12] H. Zermane, "Pengembangan Sistem Monitoring Produksi Semen yang Efisien Berbasis Algoritma Random Forest yang Disempurnakan Abstrak," pp. 1–16, 2022.

-
- [13] S. Herho and I. Dasapta Erwin, "PY-METEO-NUM: Dockerized Python Notebook Environment for Portable Data Analysis Workflows in Indonesian Atmospheric Science Communities," *Int. J. Data Sci.*, vol. 2, no. 1, pp. 38–46, 2020, doi: 10.18517/ijods.2.1.38-46.2021.
- [14] T. R. Harjanto, A. Prastya, S. Bahri, and O. Prasadi, "Analisis Kontribusi Pemanfaatan Limbah Kantong Semen (Reject) Berdasarkan Prespektif Life Cycle Assessment (Studi Kasus : PT. Solusi Bangun Indonesia Tbk.)," *J. Pengendali. Pencemaran Lingkung.*, vol. 5, no. 2, pp. 126–135, 2023, doi: 10.35970/jppl.v5i2.1900.
- [15] A. S. R. Hasibuan, W. Wahyuda, and F. D. Sitania, "Inventory management of 50 kg packaged cement products with a lot sizing ratio (case study: XYZ warehouse)," *Tek. J. Sains dan Teknol.*, vol. 17, no. 2, p. 235, 2021, doi: 10.36055/tjst.v17i2.12684.