



DATA SERIES 16.0 :

AI/ML SUMMARY & PROJECT

7 - 10 Januari 2025 

19.00 - 21.00 WIB 

Data Series 16.0 - 2025

Presented by : Evriza Ramadhan Maulana

 <https://www.linkedin.com/in/evriza-ramadhan-maulana>

 github.com/Adhonevriza  adhonevriza@gmail.com

LET'S TALK ABOUT ARTIFICIAL INTELLIGENCE (AI) & MACHINE LEARNING (ML)

- Pengertian dibaliknya
- Bagaimana Aplikasinya
- Jenis-Jenis Machine Learning
- Istilah-Istilah dalam Machine Learning

Underfit

Overfit

Goodfit



WHAT IS AI/ML?

01

• *Artificial Intelligence (AI)*

Artificial Intelligence merupakan sistem yang meniru kecerdasan manusia, sehingga dapat merasakan (sense), menalar (reason), bertindak (act), dan beradaptasi (adapt). AI menggunakan algoritma dan model matematika untuk memahami, belajar, dan mengambil keputusan berdasarkan data.

• *Machine Learning (ML)*

Machine Learning merupakan sub-bidang yang berfokus untuk pengembangan algoritma yang belajar dari data, sehingga kinerjanya meningkat seiring dengan semakin banyaknya data yang diterima seiring waktu berjalan.

• *Applications of Machine Learning*

- Virtual Personal Assistant
- Image, Text, & Speech Recognition
- Email Spam & Malware Filtering
- Self-driving Cars
- Automatic Language Translation
- Traffic Prediction



CATEGORY OF MACHINE LEARNING

1. Supervised Learning

Supervised Learning merupakan jenis metode pembelajaran mesin dimana hasil yang diharapkan pengguna sudah diketahui atau dimiliki informasinya oleh sistem. Hal ini berarti bahwa metode pembelajaran ini bekerja dengan memanfaatkan kembali data-data dan hasil output yang pernah dimasukkan oleh pengguna atau dikerjakan oleh sistem sebelumnya.

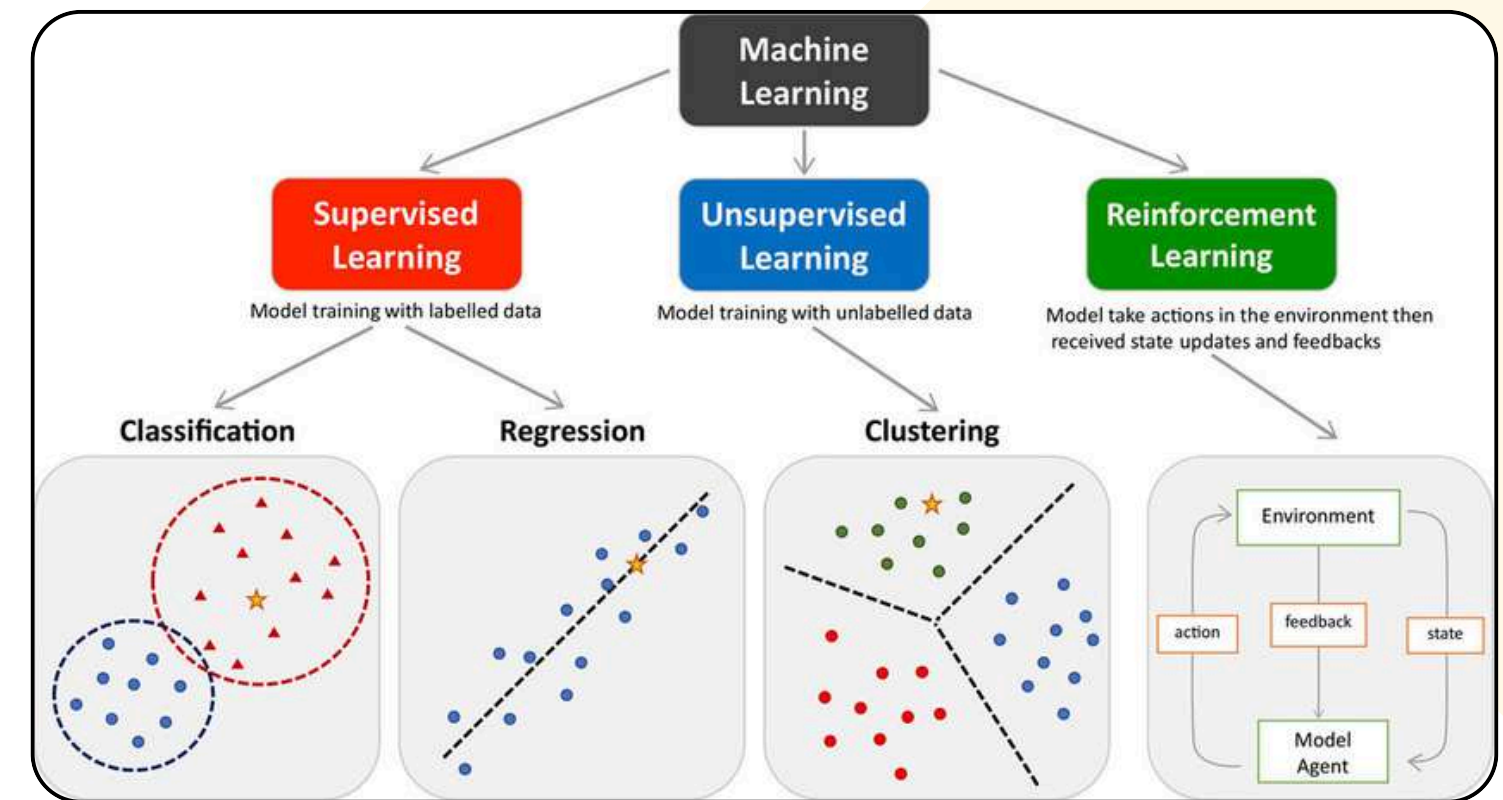
2. Unsupervised Learning

Unsupervised Learning merupakan jenis metode pembelajaran mesin dimana hasil yang diharapkan tidak dapat diketahui oleh siapapun. Dengan kata lain, hasil yang akan ditampilkan hanya bergantung kepada nilai bobot yang disusun pada awal pembangunan sistem. Tujuan utama dari metode pembelajaran ini adalah agar para penggunanya dapat mengelompokkan objek-objek yang dinilai sejenis dalam ruang atau area tertentu.

3. Reinforcement Learning

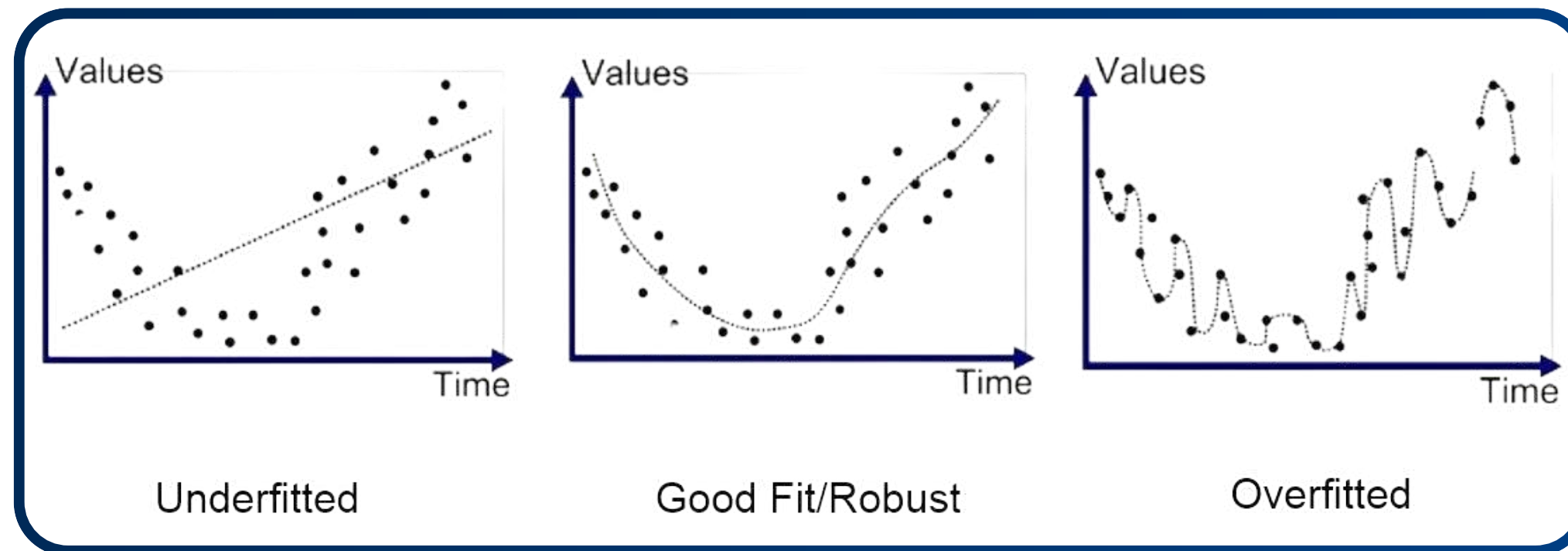
Reinforcement Learning merupakan jenis metode pembelajaran mesin dengan model yang belajar menggunakan sistem reward dan punishment. Mesin akan mempelajari bagaimana cara untuk membuat keputusan terbaik secara berurutan, untuk memaksimalkan ukuran sukses kehidupan nyata

Jenis-Jenis Machine Learning



UNDERFITTING & OVERFITTING IN MACHINE LEARNING

03



Underfit

Goodfit

Overfit

Dalam machine learning, kinerja model dievaluasi berdasarkan akurasi dan generalisasi. Akurasi berarti seberapa baik model memprediksi nilai target yang tepat, dan generalisasi berarti seberapa baik model berperilaku pada data yang terlihat dan data tidak terlihat.

Underfit

Terjadi saat model gagal belajar, sehingga nilai akurasi data training & data testing rendah

Overfit

Terjadi saat model terlalu menyesuaikan data training sehingga nilai akurasi data testing rendah, namun nilai akurasi di data training tinggi

Goodfit

Terjadi saat model belajar cukup dari data, sehingga nilai akurasi pada data testing dan data training sama-sama tinggi.

PROJECT

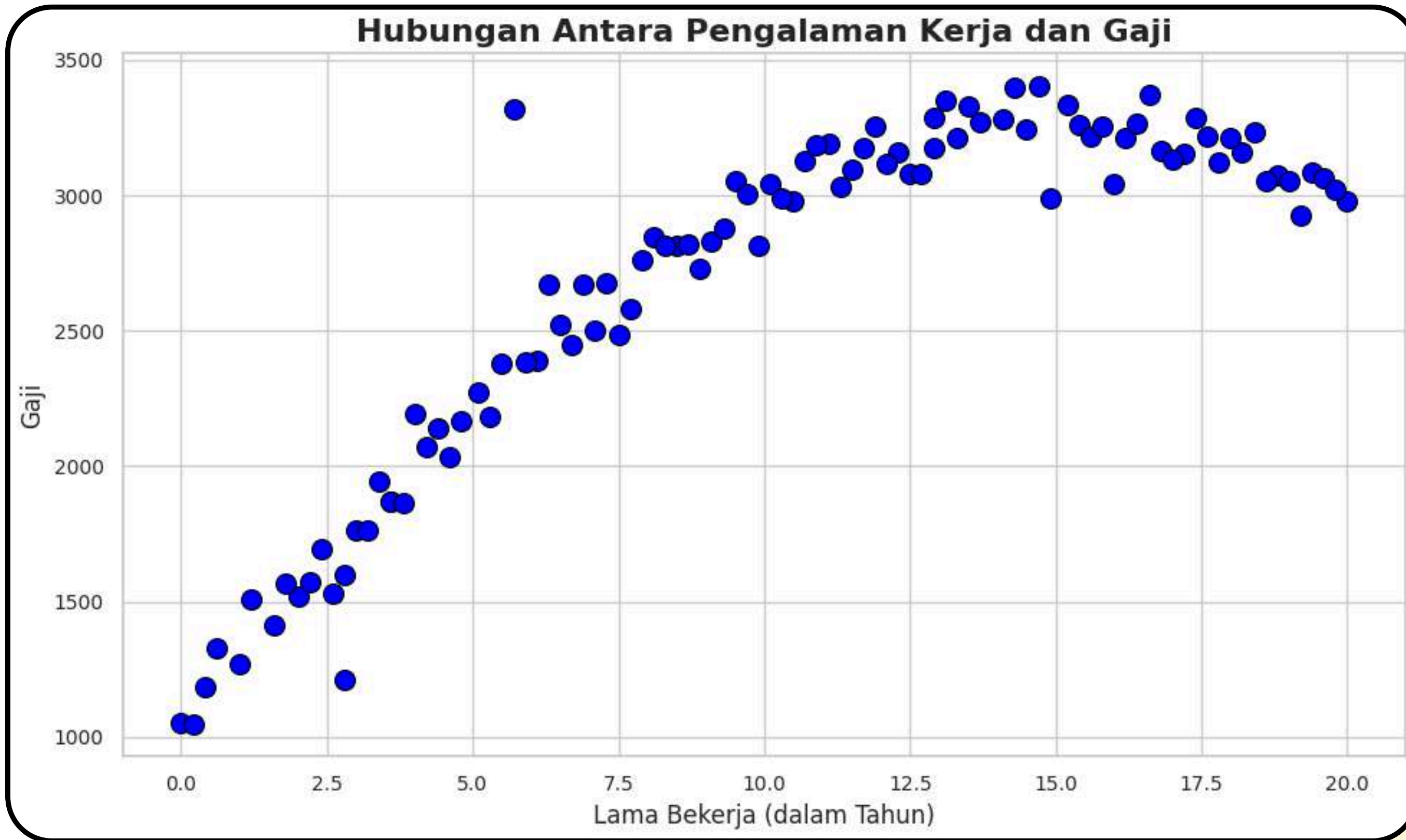
SUPERVISED - SALARY PREDICTION

Linear Regression, Decision Tree, and
Random Forest Model using Google Colab

- Data : https://github.com/Adhonevriza/best-salary-prediction-model/blob/main/salary_data.csv
- Google Colab : https://github.com/Adhonevriza/best-salary-prediction-model/blob/main/Salary_Prediction.ipynb

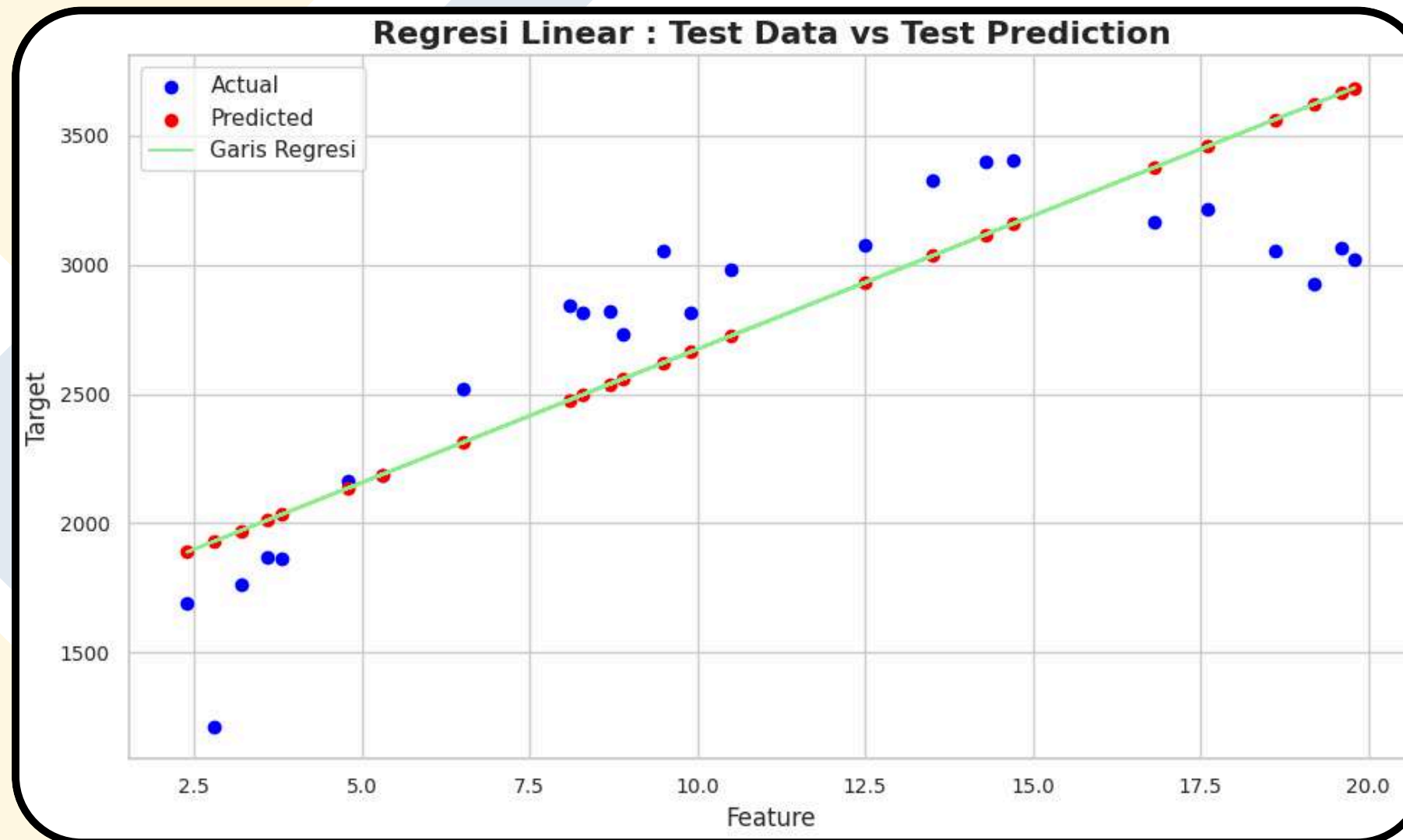


SUPERVISED - SALARY PREDICTION



SUPERVISED - SALARY PREDICTION

LINEAR REGRESSION MODEL



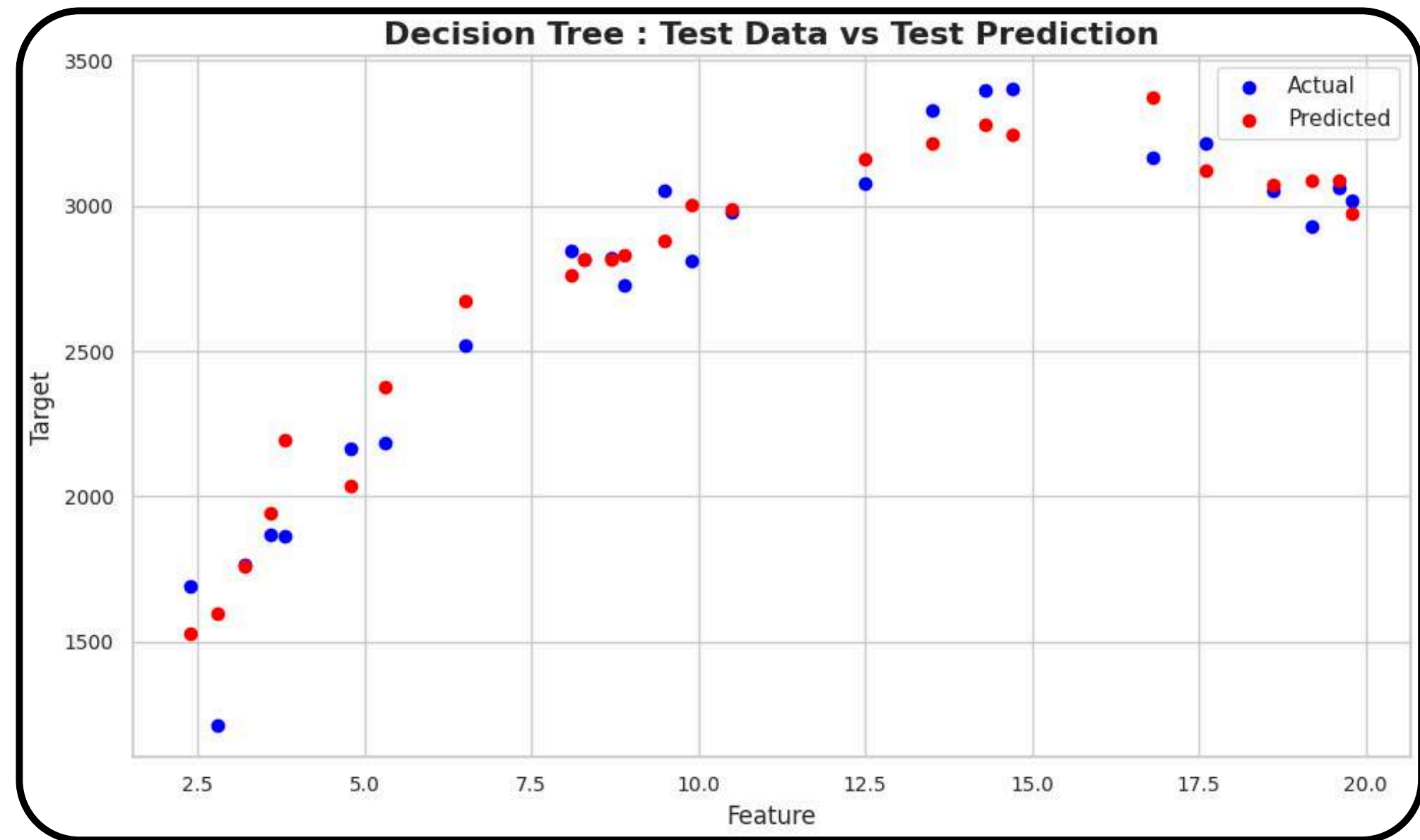
Mean Squared Error (MSE) :
Train : 107699.85
Test : 128111.12
Gap : 20411.27
Nilai R^2 :
Train : 0.77
Test : 0.63

Intercept model Regresi Linear : 1641.365997520288
Koefisien garis Regresi Linear : [103.19725286]

Maka, model Regresi Linearnya adalah
 $y = 1641.366 + 103.197 x$

SUPERVISED - SALARY PREDICTION

DECISION TREE MODEL



Mean Squared Error (MSE) :

Train : 88.12

Test : 23627.99

Gap : 23539.87

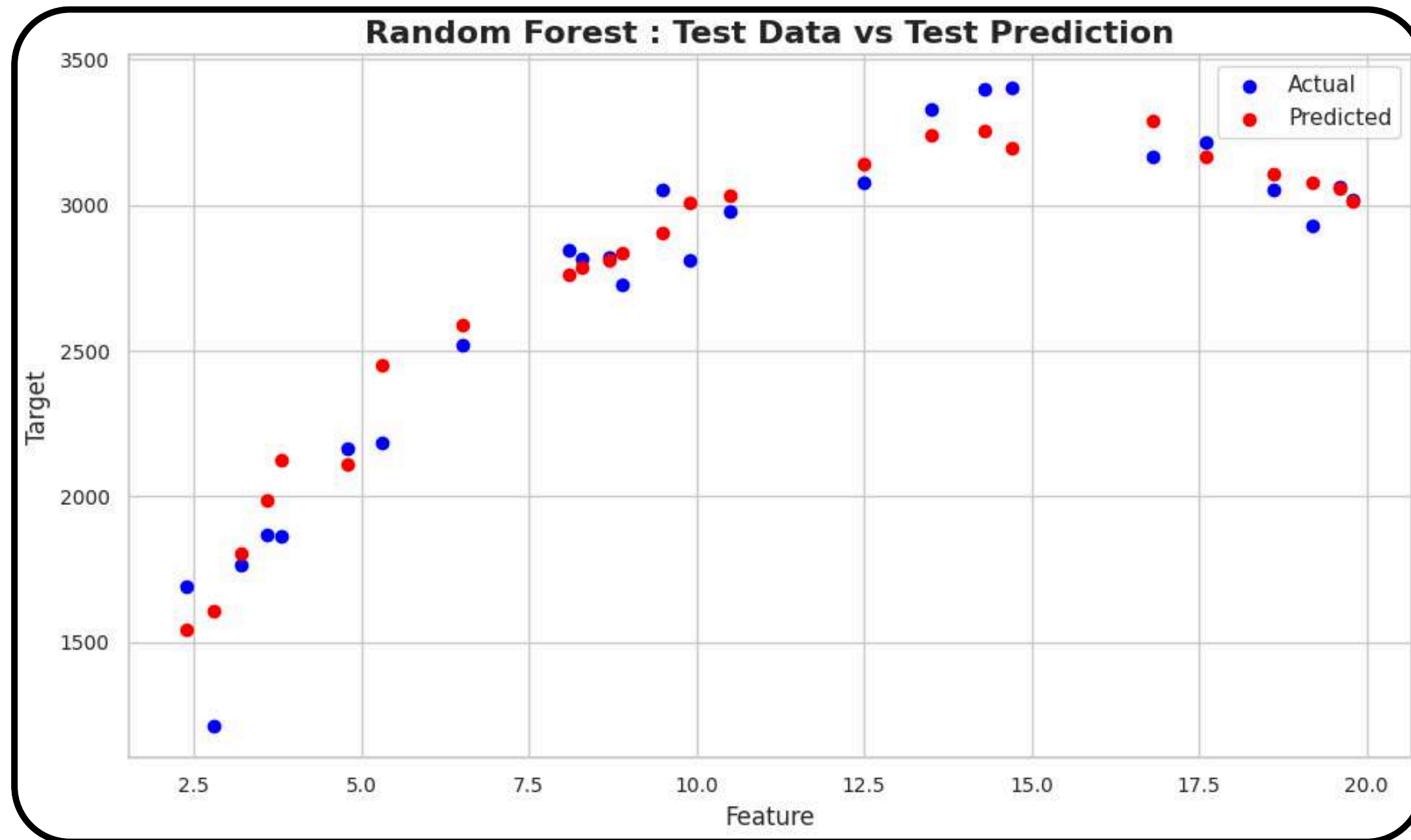
Nilai R^2 :

Train : 1.00

Test : 0.93

SUPERVISED - SALARY PREDICTION

RANDOM FOREST MODEL



```
Mean Squared Error (MSE) :  
Train : 3737.44  
Test  : 21744.73  
Gap   : 18007.29  
Nilai R^2 :  
Train : 0.99  
Test  : 0.94
```

SUPERVISED - SALARY PREDICTION

```
Mean Squared Error (MSE) :  
Train : 107699.85  
Test : 128111.12  
Gap : 20411.27  
Nilai R^2 :  
Train : 0.77  
Test : 0.63
```

Regresi Linear

Memiliki selisih nilai MSE yang cukup besar dengan nilai koefisien determinasi yang cukup kecil

```
Mean Squared Error (MSE) :  
Train : 88.12  
Test : 23627.99  
Gap : 23539.87  
Nilai R^2 :  
Train : 1.00  
Test : 0.93
```

Decision Tree

Memiliki selisih nilai MSE terkecil dengan nilai koefisien determinasi yang sangat baik

```
Mean Squared Error (MSE) :  
Train : 3737.44  
Test : 21744.73  
Gap : 18007.29  
Nilai R^2 :  
Train : 0.99  
Test : 0.94
```

Random Forest

Memiliki selisih nilai MSE terbesar dengan nilai koefisien determinasi yang sangat baik

CONCLUSION

Di antara ketiga model yang telah diujikan, **model Random Forest merupakan model terbaik**. Penilaian ini didasarkan pada beberapa indikator :

1. Mean Squared Error (MSE) :

- Train : 3737,44 (terendah, menunjukkan tingkat error kecil pada data training)
- Test : 21744,73 (lebih kecil daripada model lain, menunjukkan performa yang lebih baik pada data test)

2. Gap / Selisih antara MSE Train dan MSE Test :

- Model Random Forest memiliki gap / selisih terkecil (18007,29), menunjukkan bahwa hasil prediksi menggunakan model ini sesuai dengan data aktual dan lebih baik dibandingkan Regresi Linear (20411,27) dan Decision Tree (23539,87).

3. Nilai Koefisien Determinasi (R kuadrat) :

Koefisien Determinasi merupakan indikator yang mengukur seberapa baik model sesuai dengan data dan seberapa baik model dapat memprediksi hasil di masa mendatang. Adapun nilai Koefisien Determinasi dari model ini adalah sebagai berikut :

- Train : 0,99 (sangat baik)
- Test : 0,94 (tertinggi di antara ketiga model)

Secara keseluruhan, model Random Forest memiliki keseimbangan antara performa pada data training dan data testing, dengan nilai MSE yang kecil dan nilai koefisien determinasi yang tinggi. Model ini merupakan pilihan terbaik di antara ketiga model untuk konteks ini.



Data Series 16.0 - AI/ML

Thank You



Contact Me :

 <https://www.linkedin.com/in/evriza-ramadhan-maulana>

 github.com/Adhonevriza  adhonevriza@gmail.com

