AI & Machine Learning

Salary Prediction According Years of Work Experience

Raya Ahmad Syarif

Artificial intelligence is a system that mimics how humans think. AI works by utilizing data, training models and applying algorithms. AI can perform a variety of tasks given by humans.

These tasks include learning, reasoning, problem-solving, perception, and language understanding.

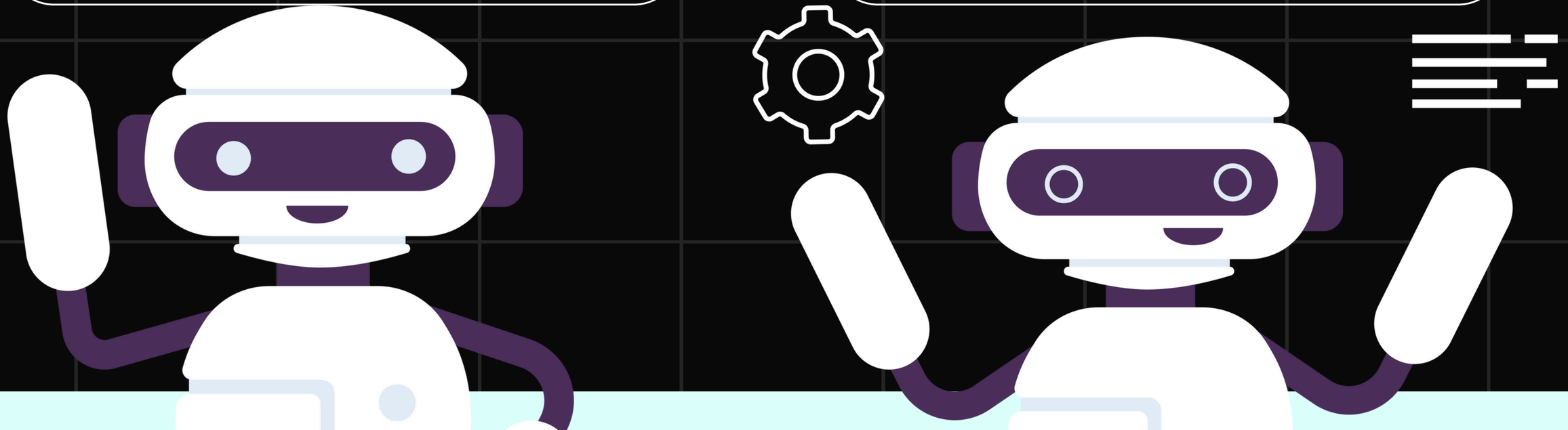
What is Machine Learning & Deep Learning?

Machine Learning

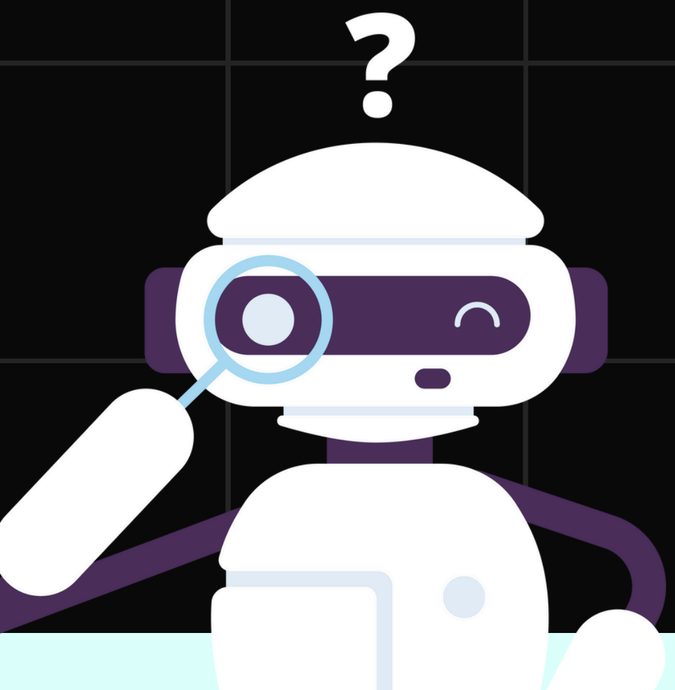
Subfield of AI that can learn and adapt without given instructions. But by utilizing data.

Deep Learning

Subset of machine learning using neural networks with many layers, capable of learning representations of data.



Types of Machine Learning



Reinforcement Learning

A machine learning technique that focus on decision making to achieve the best outcome. It uses reward-and-punishment to get the best decision.

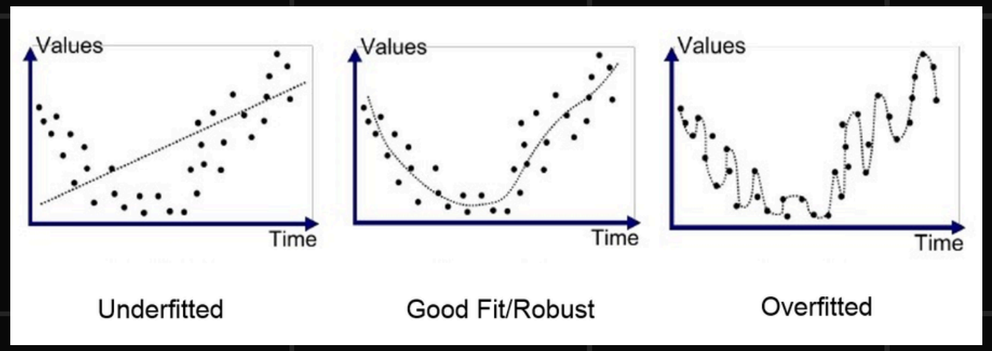
Supervised Learning

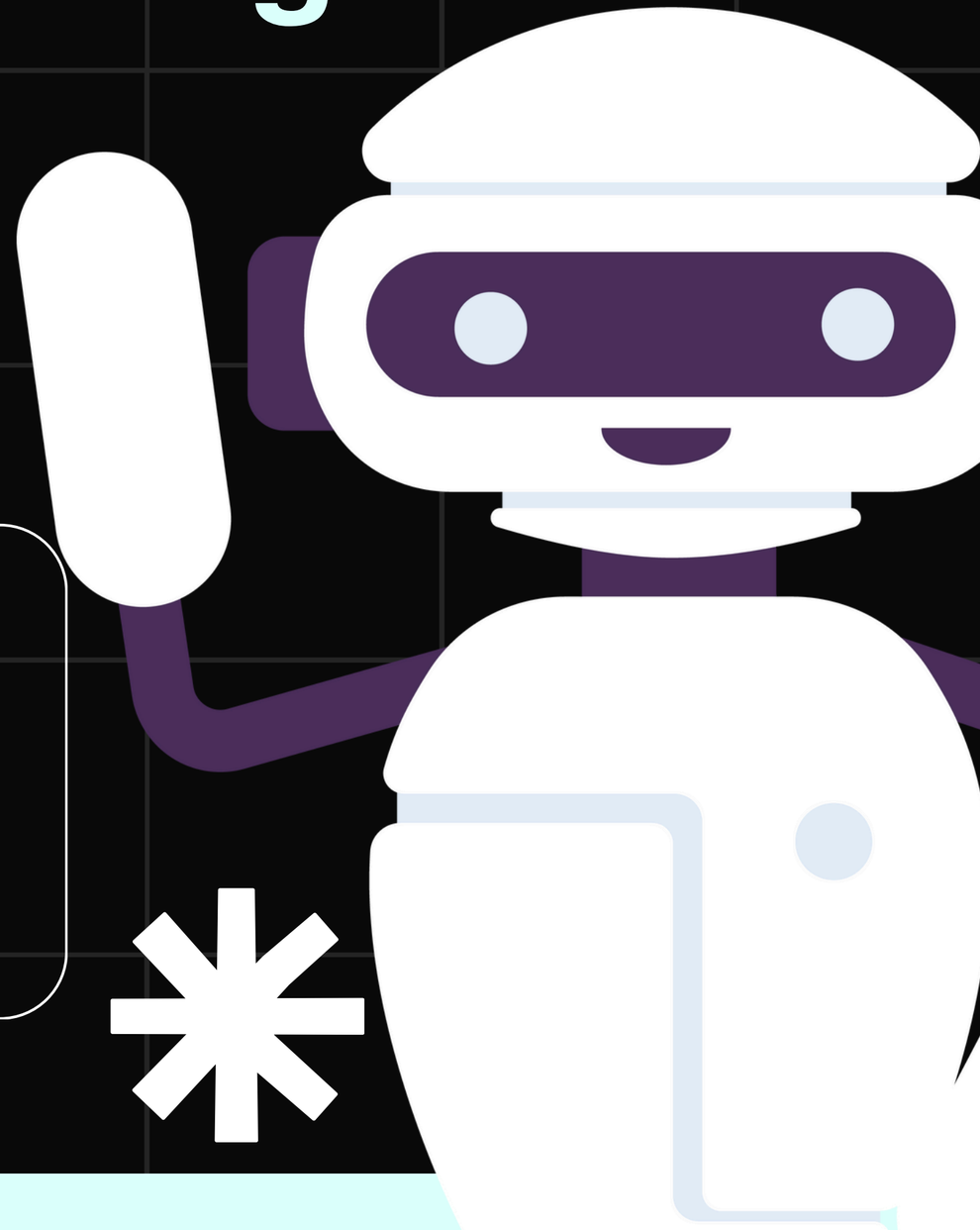
A machine learning techniques that uses labeled data to predict outcomes by learning patterns.

Unsupervised Learning

Instead using labeled data it uses unlabeled data without any human interventions. It can find patterns and relationships in data.

Overfitting and Underfitting?





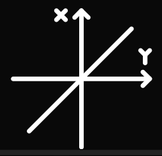
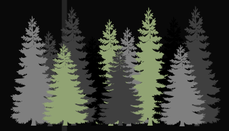
Underfitting

Underfitting is a condition where the machine learning model can’t learn the relation between variables in data. Because of this the ML Model can’t predict outcomes.

Overfitting

Overfitting occurs when a machine learning model learns the training data too well, including the noise and irrelevant details. This leads to a failure in generalizing to new data, meaning the model struggles to make accurate predictions on unseen inputs.

Analysis Procedure

Analysis	Description		
Exploratory Data Analysis	 Data Description	 Data Cleaning	 Data Splitting
Machine Learning and Modelling	 Linear Regression	 Decision Tree	 Random Forest

Data Description

Columns and Target Data

There are three columns and one target data.

Number of Samples

100 entries consist of 0 missing value and 3 duplicates.

Description:

Salary increases with years of work experience, but starts to decline after 15 years work experience

Removing duplicates

There are two duplicate datas that needs to be removed.

Data Cleaning

	employee_id	experience_years	salary
10	EM_111	3.6	1867.9
42	EM_111	3.6	1867.9

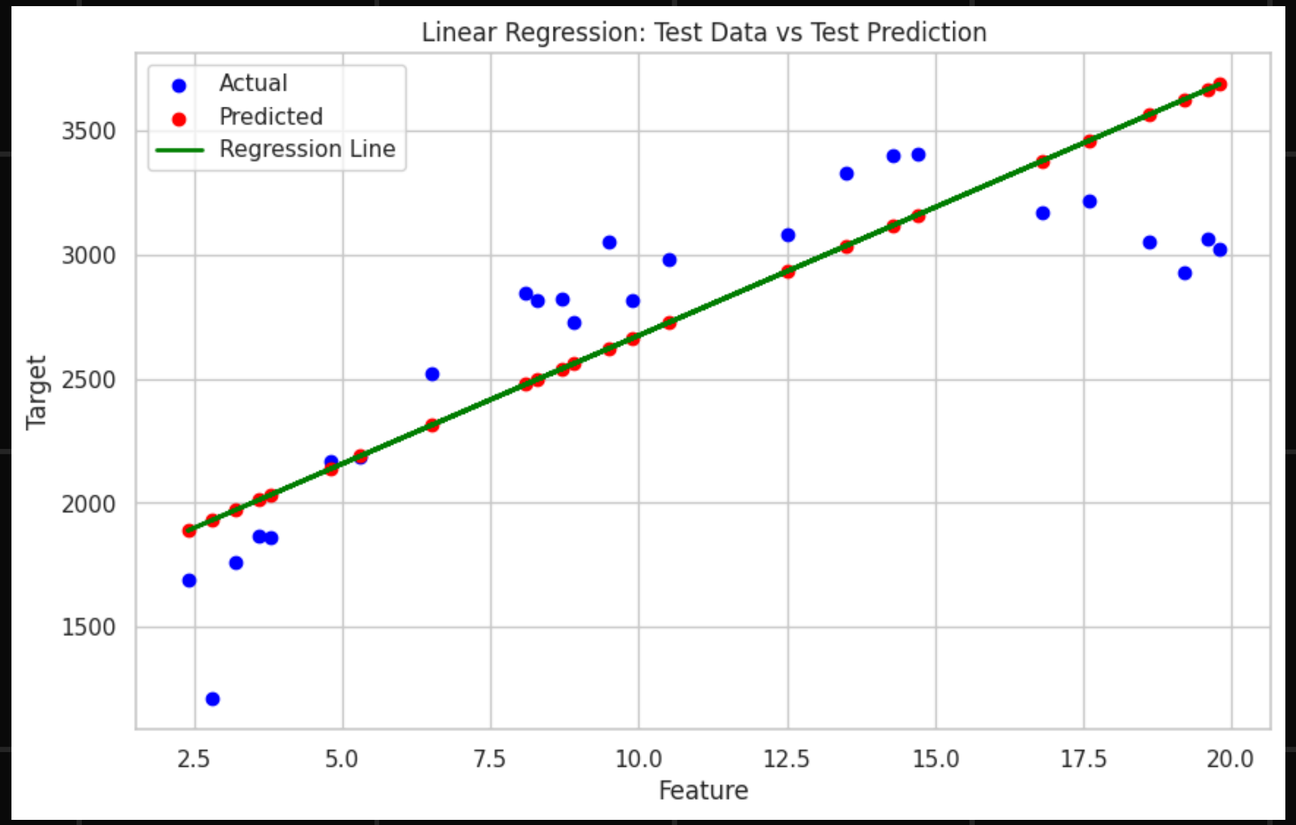
Splitting data by Columns

We split the two columns which is experience years and salary.

Data Splitting

	salary
0	3166.9
1	3126.9
2	3278.8
3	2828.8
4	2728.7

	experience_years
0	16.8
1	10.7
2	14.1
3	9.1
4	8.9



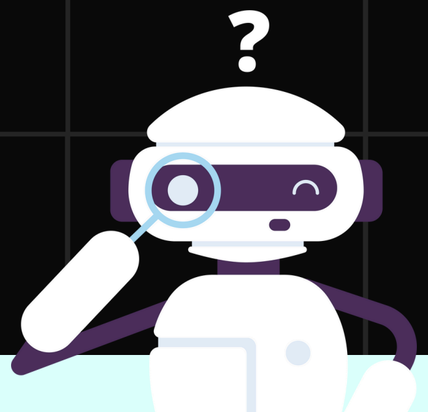
Output:

Mean Squared Error:
Train: 107699.85
Test : 128111.12
Gap : 20411.27

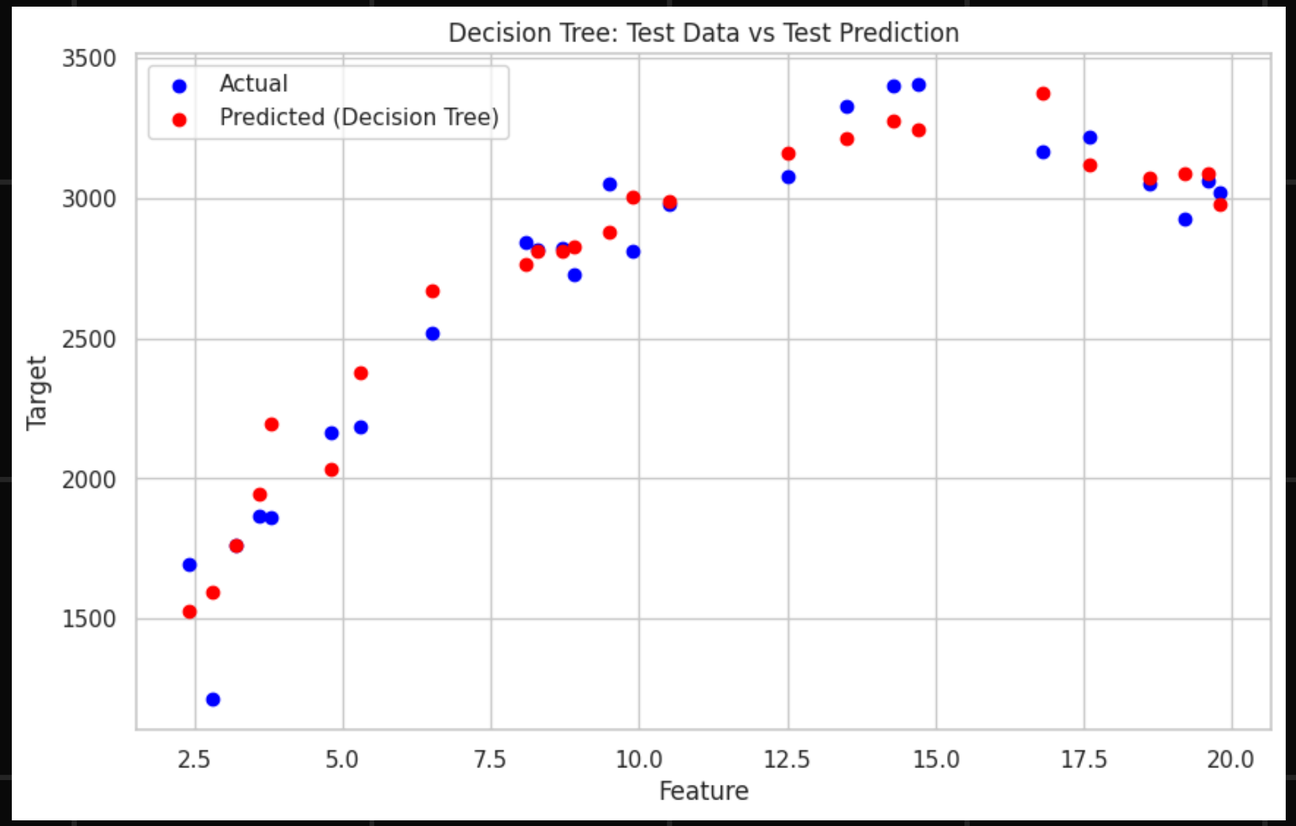
R^2 Score:
Train: 0.77
Test : 0.63

Linear Regression Model:

y = 1641.366 + 103.197 x



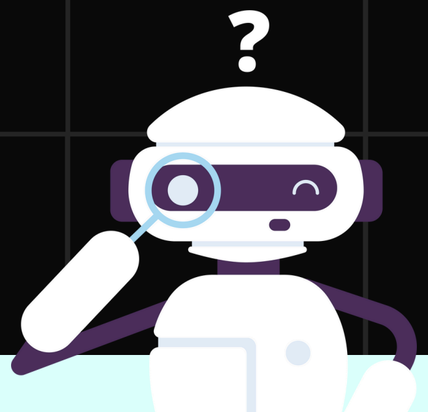
Linear Regression



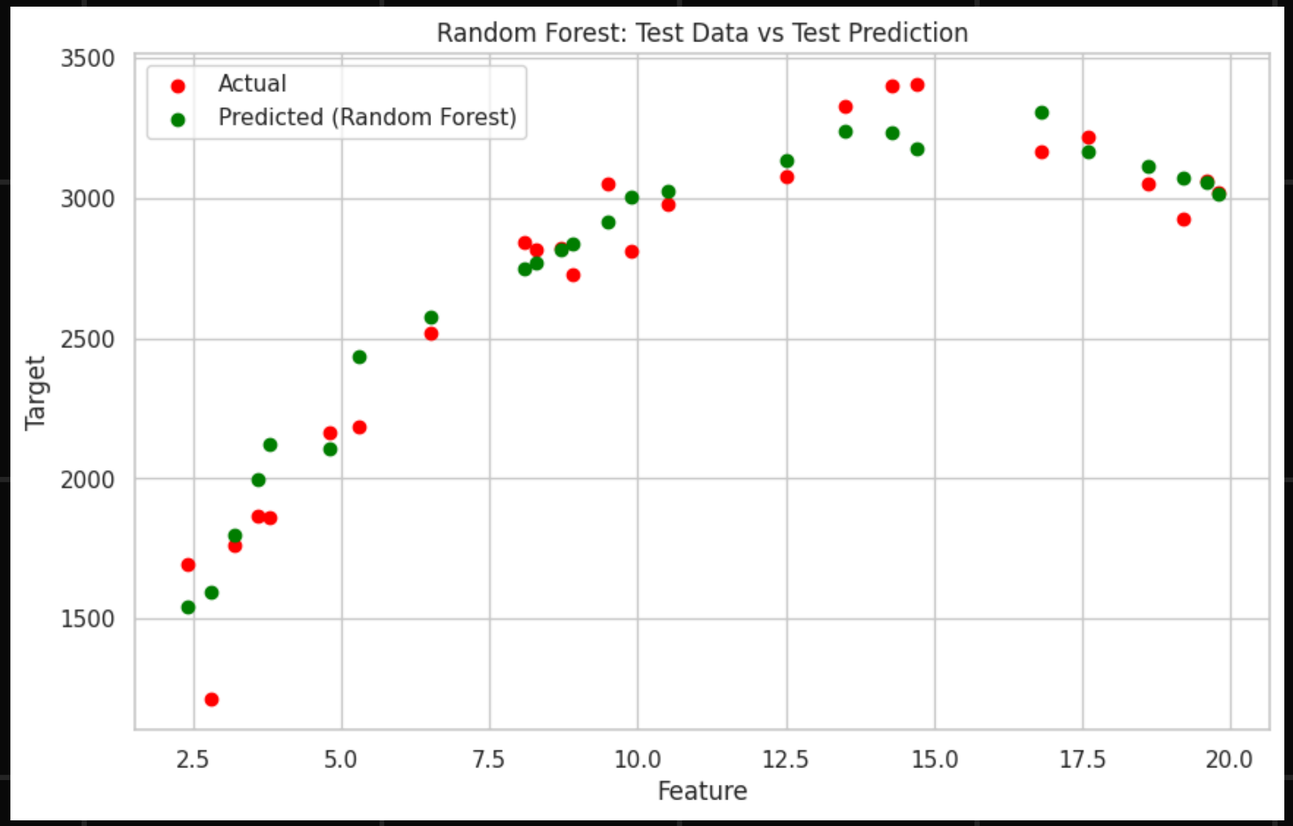
Output:

Mean Squared Error:
Train: 88.12
Test : 23627.99
Gap : 23539.87

R^2 Score:
Train: 1.00
Test : 0.93



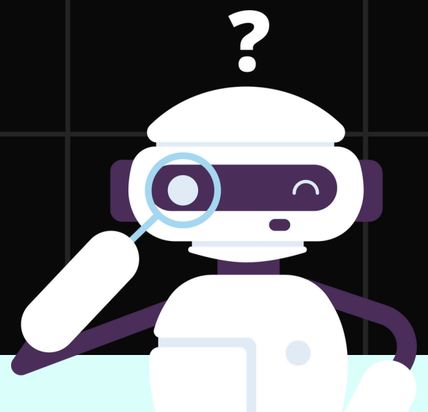
Decision Tree



Output:

Mean Squared Error:
Train: 3677.29
Test : 21628.63
Gap : 17951.35

R^2 Score:
Train: 0.99
Test : 0.94



Random Forest

**

Thank you