

SALARY PREDICTION

AI & MACHINE LEARNING

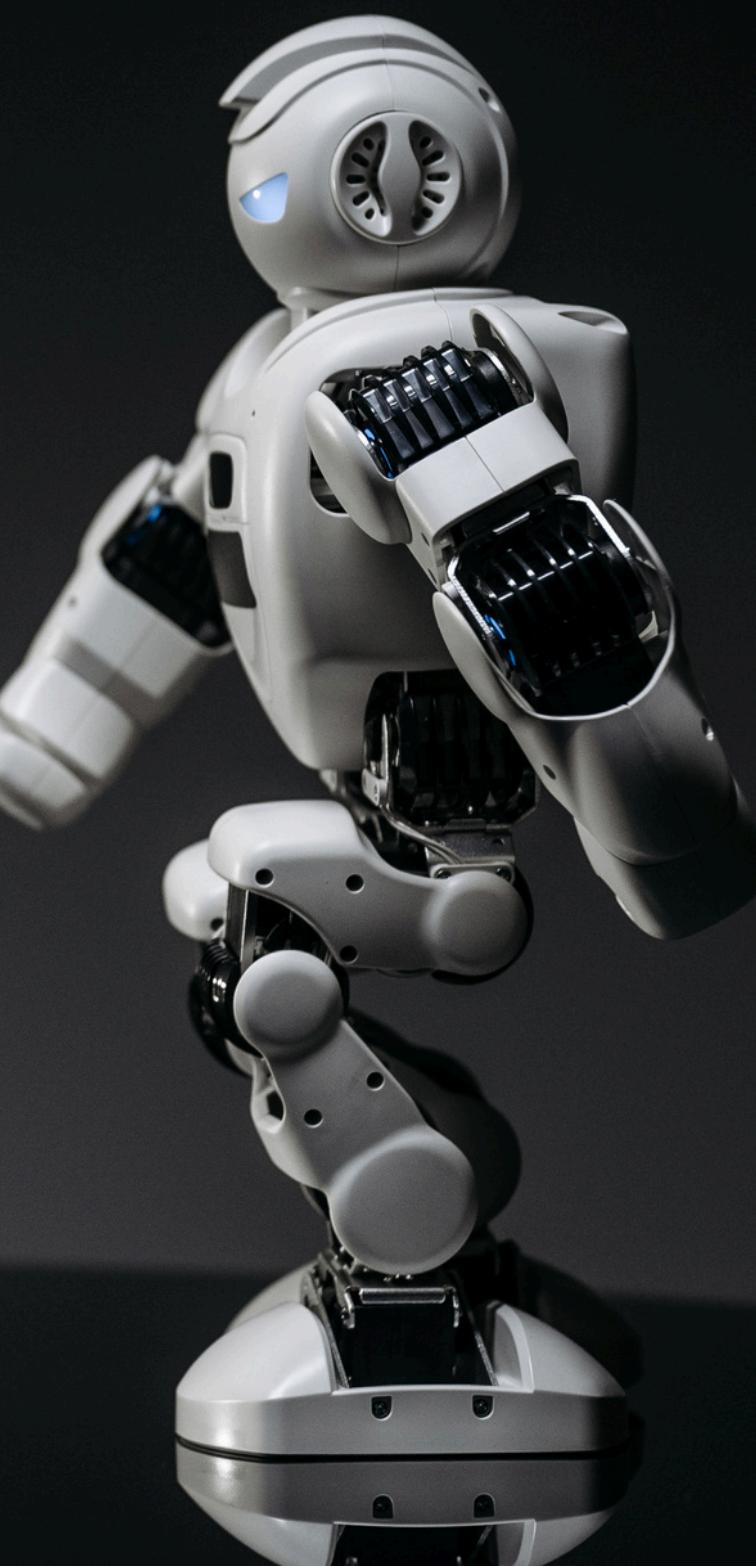
Muhammad Ahsani Taqwim Rakhman

@reallygreatsite

ARTIFICIAL INTELLIGENCE

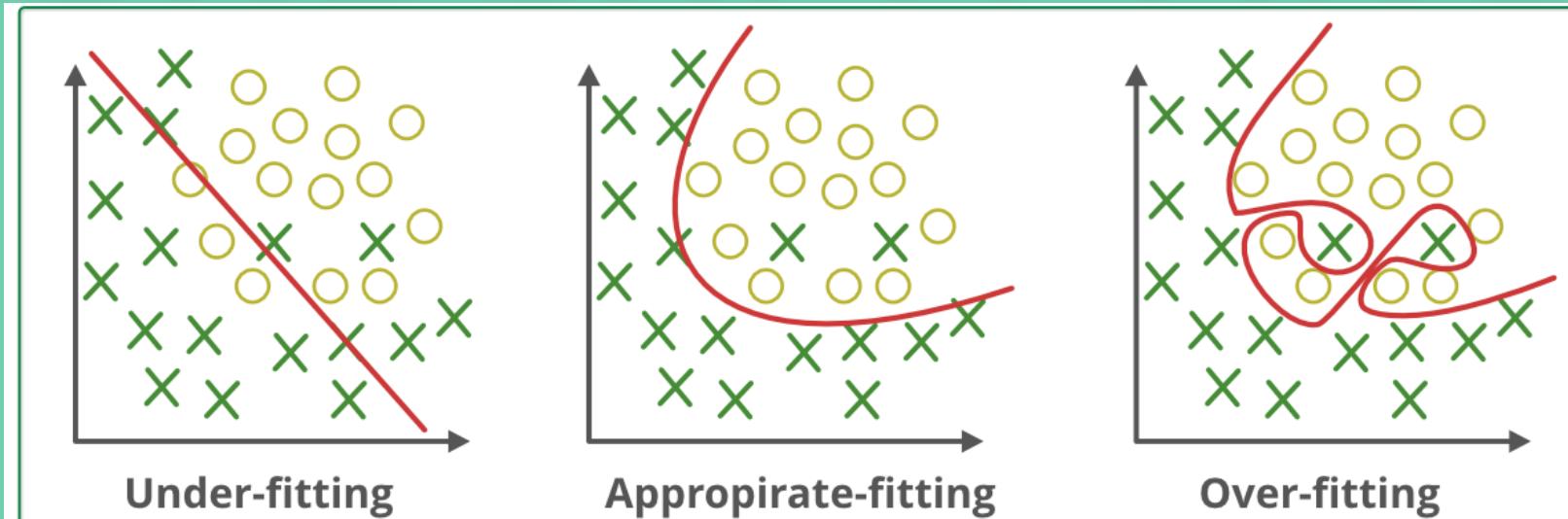
Artificial Intelligence (AI) is the broader field focused on creating computer systems that can perform tasks that typically require human intelligence. This includes things like understanding language, recognizing images, making decisions, and solving problems.

Machine Learning (ML) is a subset of AI that focuses on developing systems that can "learn" from data. Instead of following explicitly programmed rules, ML systems identify patterns in data to make predictions or decisions.



OVERFITTING AND UNDERFITTING

Slide 03



Overfitting is one of fundamental challenge in machine learning.

Overfitting occurs when a model learns the training data too perfectly, including all its noise and random fluctuations. Think of it like memorizing rather than understanding.

Signs of overfitting:

- Very high accuracy on training data
- Much lower accuracy on test data
- Model is unnecessarily complex

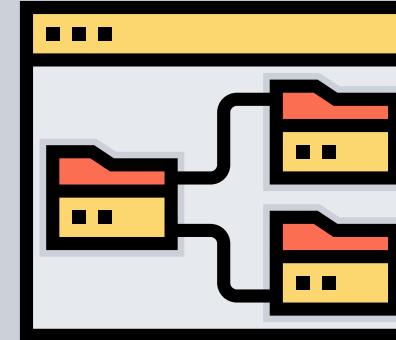
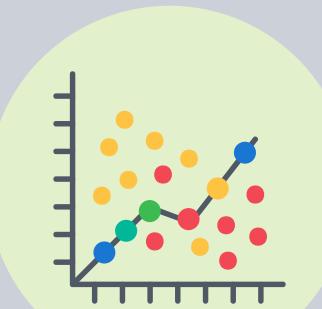
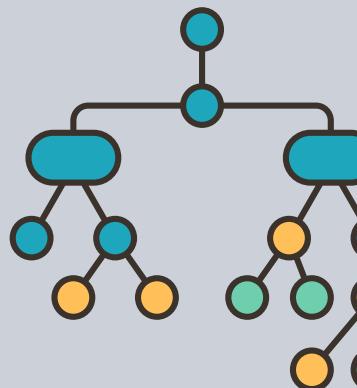
underfitting is one of fundamental challenge in machine learning.

Underfitting happens when a model is too simple to capture the underlying patterns in the data

Signs of underfitting:

- Poor accuracy on training data
- Poor accuracy on test data
- Model is too simplistic for the problem

ANALYSIS PROCEDURE

Section	Procedure		
exploratory data analysis	 Data Description	 data cleaning	 data splitting
machine learning models	 linear progression	 Decision tree	 Random forest regression

DATA DESCRIPTION

- there are three columns on the data
- zero missing value
- one duplicate

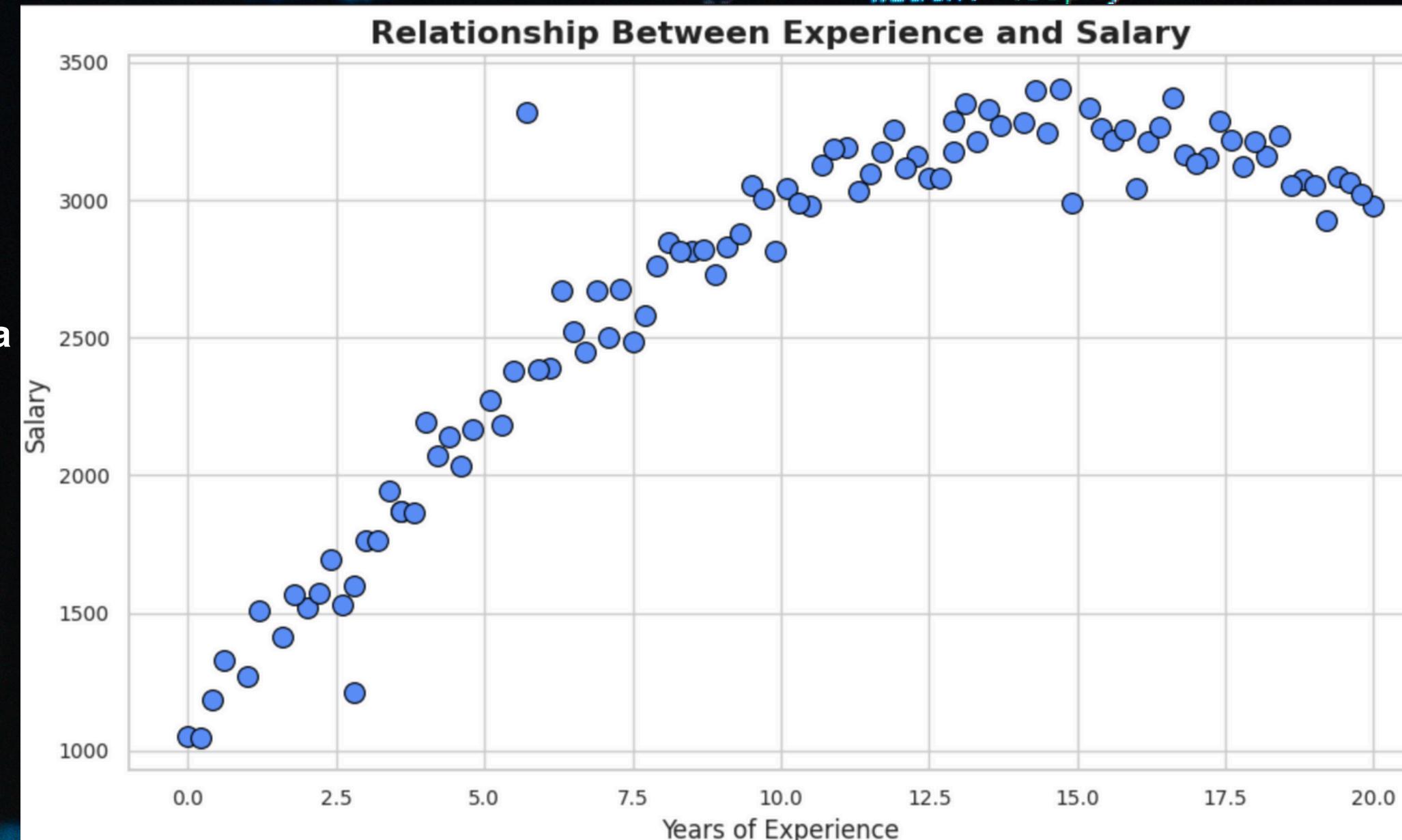
Pattern :

- From 0-10 years: There's a steep, almost linear increase in salary
- From 10-20 years: The growth rate slows down, showing more of a plateau effect
- After 15 years: There's a slight downward trend visible in the data points

Data Distribution:

- The data points are fairly evenly distributed along the experience range
- One notable outlier appears around the 5-6 year mark, showing a higher salary than the trend would predict

This pattern is typical of many professional careers, where salary growth is fastest in the early years and then tends to level off with more experience.



DATA CLEANING

- there are two duplicate data that need to be cleaned up

	employee_id	experience_years	salary
10	EM_111	3.6	1867.9
42	EM_111	3.6	1867.9

Data sebelum pemerikasaan duplikasi :
(100, 3)

before

Data setelah pemerikasaan duplikasi :
(99, 3)

after



DATA CLEANING

- there are no missing values

employee_id	0
experience_years	0
salary	0



DATA SPLITTING

splitting the data to two, turning it
into experience years (X) and salary
(y)

Intelligence

	experience_years	salary
0	16.8	3166.9
1	10.7	3126.9
2	14.1	3278.8
3	9.1	2828.8
4	8.9	2728.7

LINEAR REGRESSION

Mean Squared Error:

Train: 112597.74

Test : 108716.22

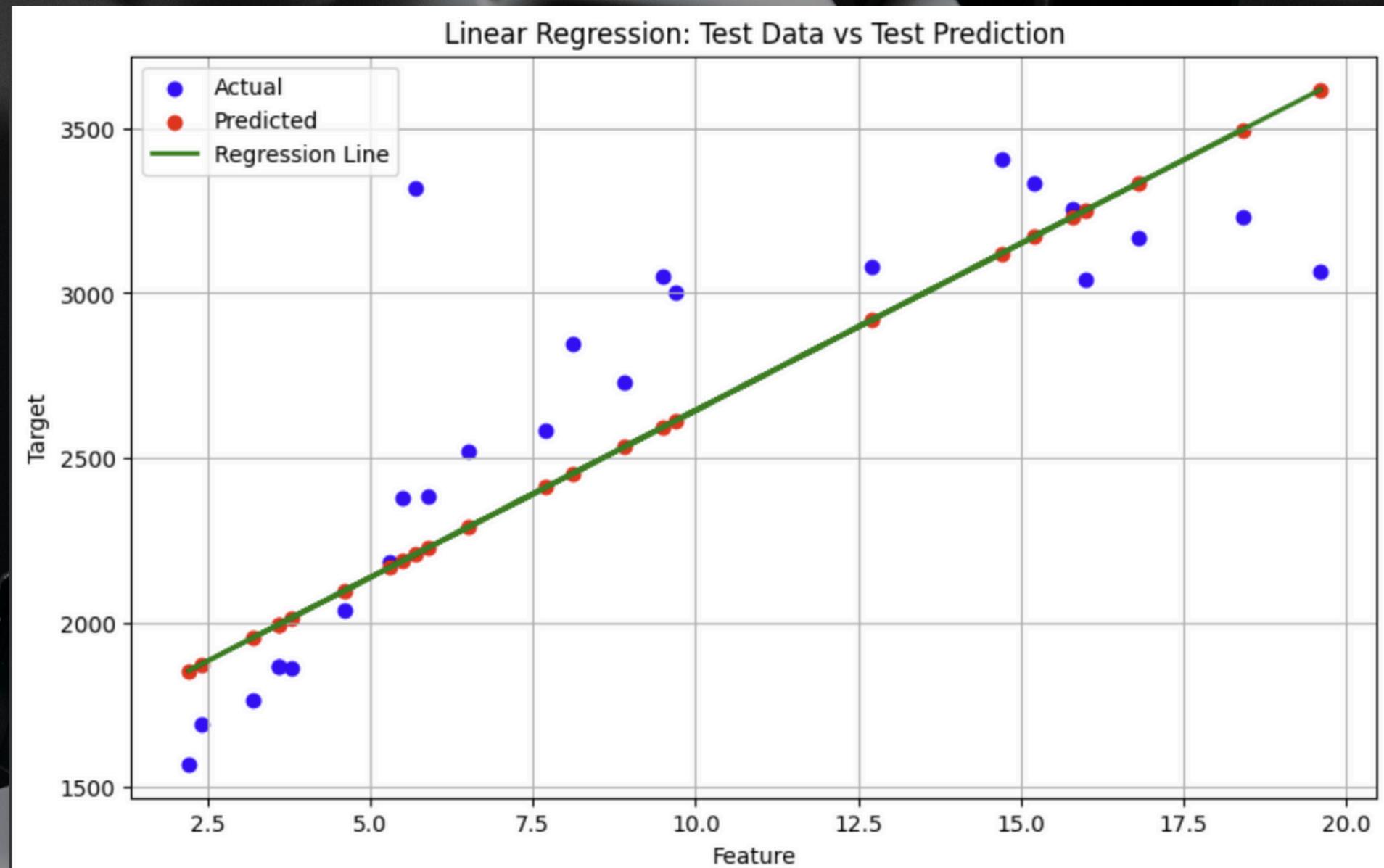
Gap : 3881.51

R² Score:

Train: 0.76

Test : 0.69

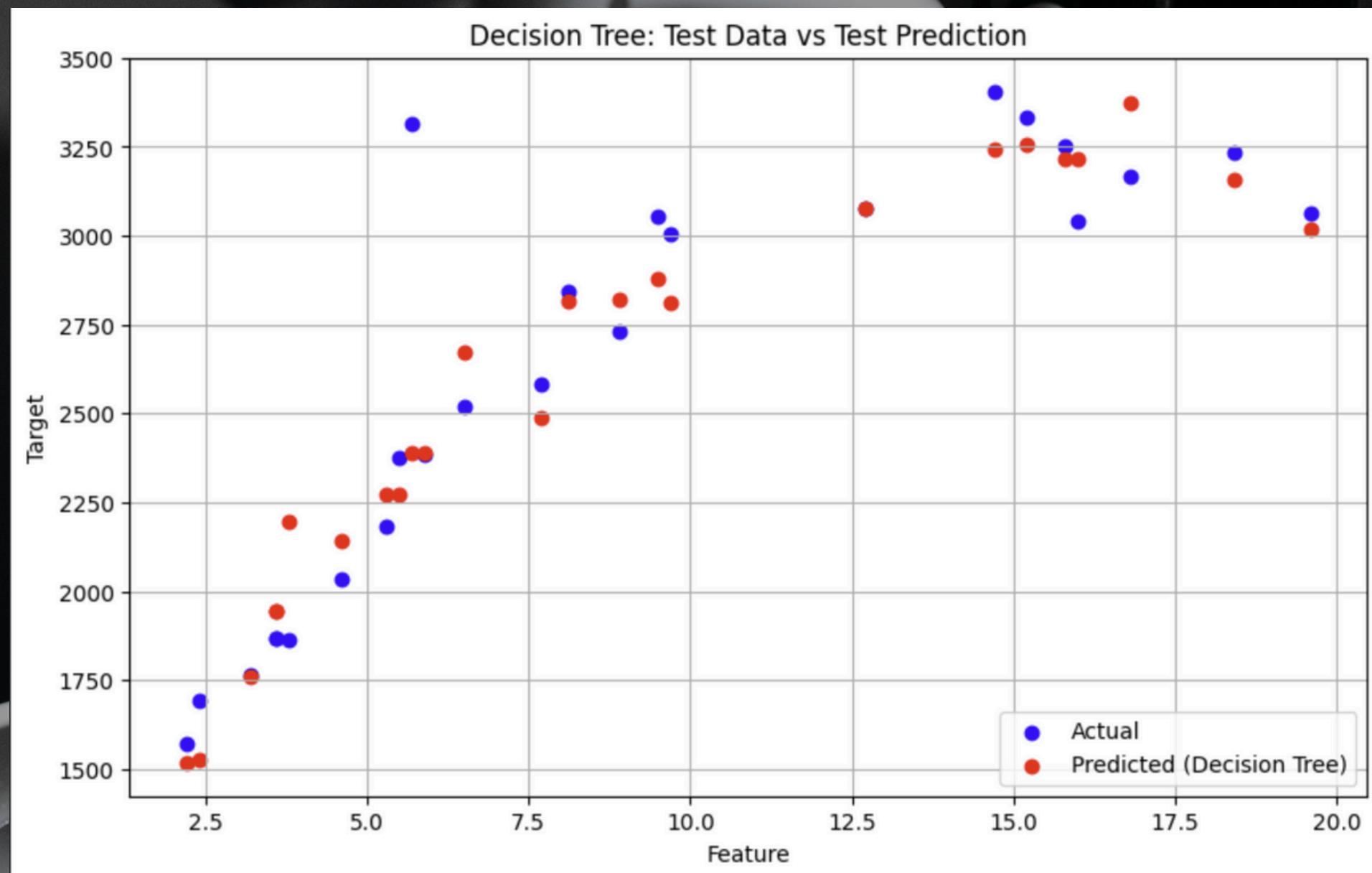
$$y = 1629.59140 + 101.42805931$$



DECISION TREE

Mean Squared Error:
Train: 1067.43
Test :50403.07
Gap : 49335.65

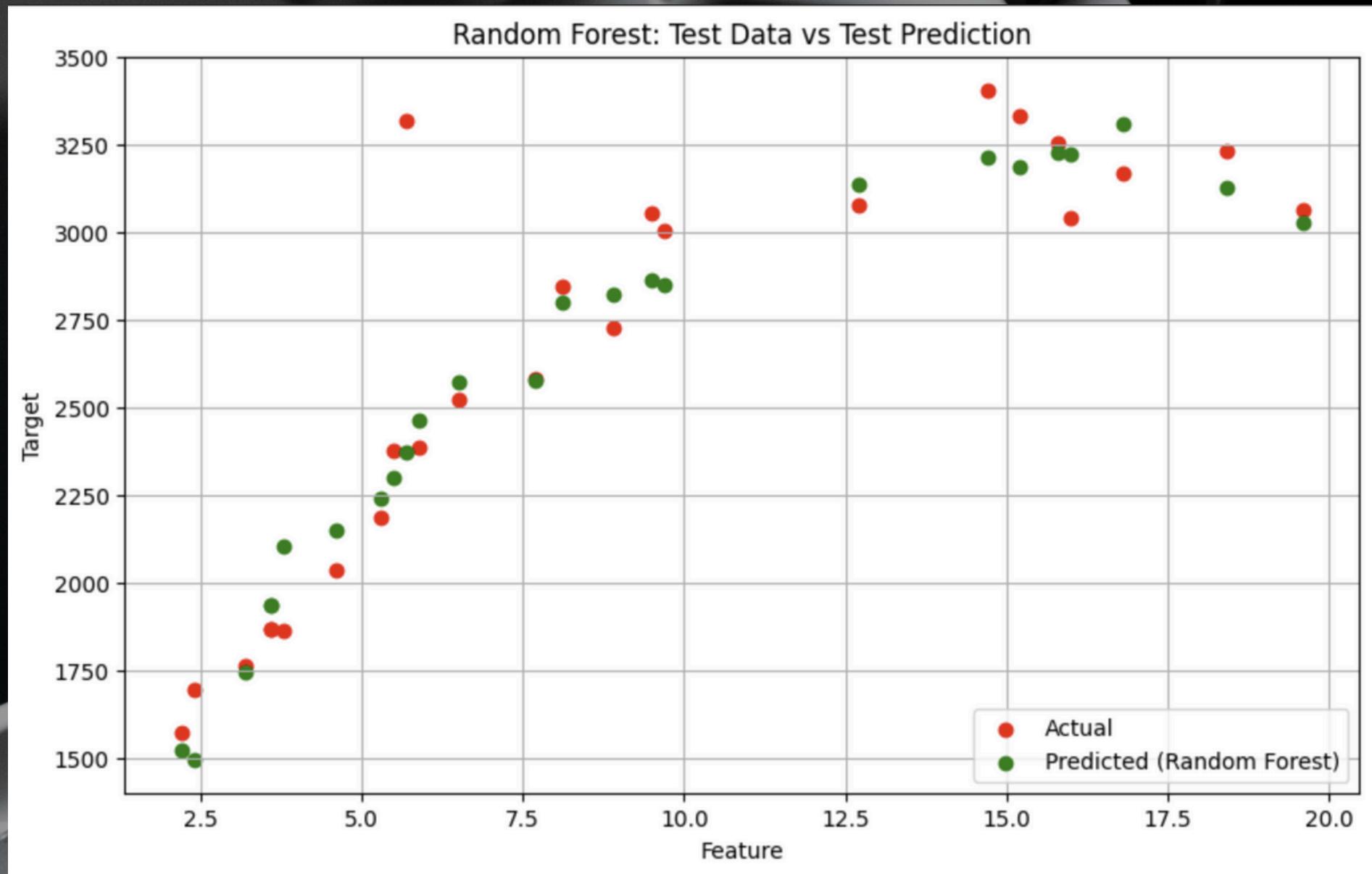
R² Score:
Train: 1.00
Test : 0.86



RANDOM FOREST

Mean Squared Error:
Train: 2908.97
Test : 49274.51
Gap : 46365.54

R² Score:
Train: 0.99
Test : 0.86





THANK YOU
Ai-machine-learning_salary-prediction.ipynb
FOR YOUR ATTENTION