AI
MACHINE LEARNING
FINAL PROJECT: SALARY PREDICTION

Designed by Ahmad Ihsan

Github

Februari-2025

CONTENT

WHAT IS AI/ML

Artificial Intelligence (AI) is the simulation of human intelligence in machines that are programmed to think and act like humans. AI enables systems to learn from data, adapt to new situations, and solve problems efficiently.

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that enables computer systems to learn from data without being explicitly programmed. ML algorithms can identify patterns in data, make predictions, and make informed decisions.

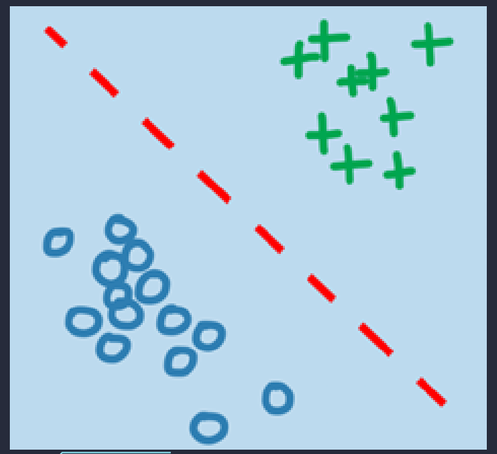
CATEGORY
MACHINE LEARNING

UNSUPERVISED LEARNING

A model that identifies patterns in data without labels.

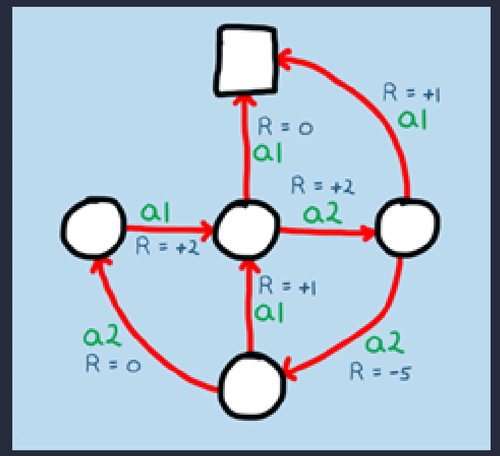
SUPERVISED LEARNING

A model that identifies patterns in labeled data.



REINFORCEMENT LEARNING

A learning model that receives rewards or punishments based on its actions.



MACHINE LEARNING PROCESS

Data

Data can be in the form of text, numbers, images, audio, or video

Preprocessing

Cleaning and preparing data to ensure its suitability for model training

Training

Training models using training data to recognize patterns or relationships within the data

Validation

Evaluating model performance during training

Testing

Evaluating the final performance of the model on unseen data

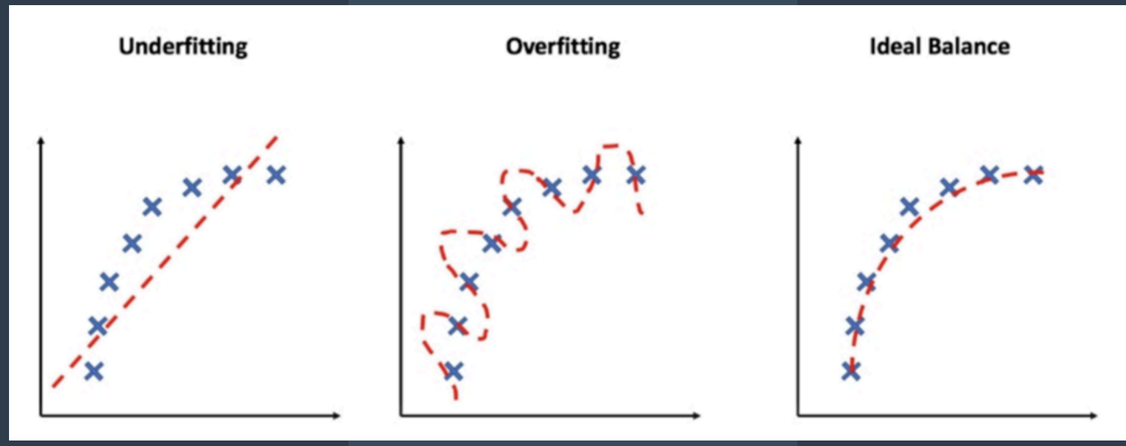
Overfit

Goodfit

Underfit

WHAT IS OVERFITTING UNDERFITTING

Overfitting occurs when a machine learning model becomes too specialized in the training data. As a result, the model becomes overly complex and fails to generalize well to new data.

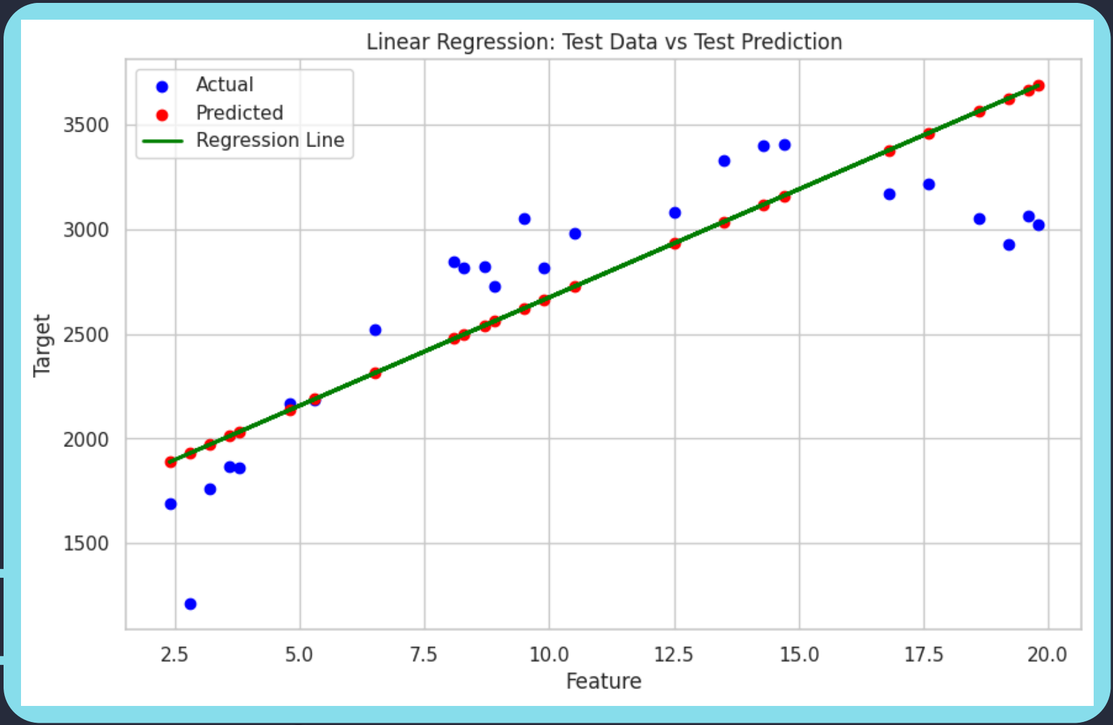


Underfitting occurs when a machine learning model is too simple to capture the underlying patterns in the data. As a result, the model performs poorly on both training data and new, unseen data.

FINAL PROJECT
SUPERVISED-SALARY PREDICTION

Check Out the Project

SALARY PREDICTION
LINEAR REGRESSION

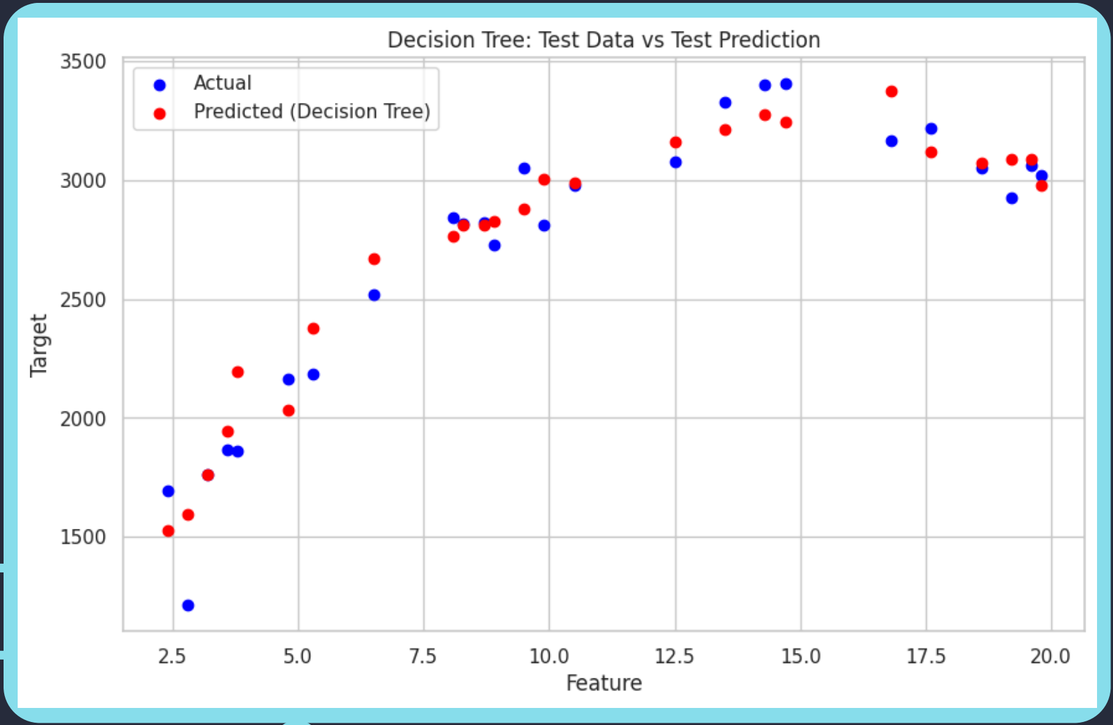


RESULT

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

R^2 Score:
Train: 0.77
Test : 0.63

SALARY PREDICTION
DECISION TREE

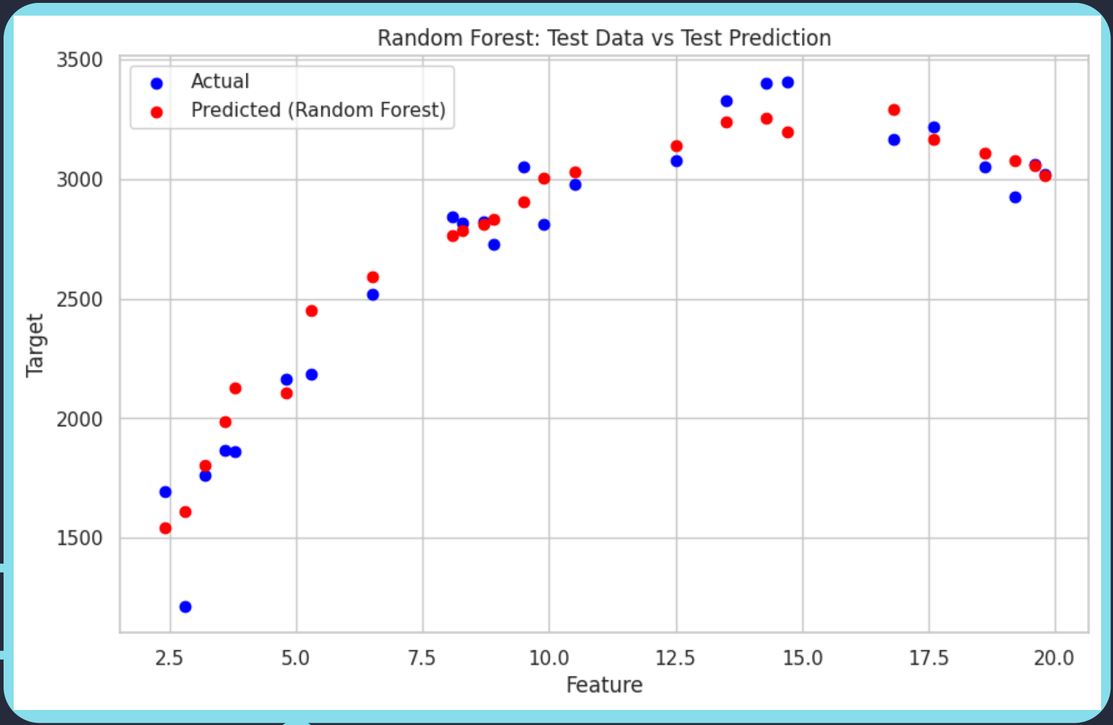


RESULT

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

R^2 Score:
Train: 1.00
Test : 0.93

SALARY PREDICTION
RANDOM FOREST



RESULT

Mean Squared Error:
Train: 3737.44
Test : 21744.73
Gap : 18007.29

R^2 Score:
Train: 0.99
Test : 0.94

COMPARISON

Linear Regression

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

R^2 Score:
Train: 0.77
Test : 0.63

Decision Tree

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

R^2 Score:
Train: 1.00
Test : 0.93

Random Forest

Mean Squared Error:
Train: 3737.44
Test : 21744.73
Gap : 18007.29

R^2 Score:
Train: 0.99
Test : 0.94

RESULTS AND CONCLUSION

Among the three tested models, Random Forest exhibited the best performance due to the following key factors:

1. High Predictive Accuracy – Random Forest achieved a strong R^2 score on both training data (0.99) and test data (0.94), demonstrating its ability to make highly accurate predictions.
2. Lower Error Rate – The Mean Squared Error (MSE) on the test data was significantly lower (21,744.73) compared to the Decision Tree model, ensuring more precise predictions.
3. Optimal Model Balance – With a minimal error gap between training and test data (18,007.29), Random Forest effectively avoids overfitting (as seen in Decision Tree) and underfitting (as observed in Linear Regression), making it a well-balanced and reliable model.

THANK YOU

Feel free to reach out if you'd like to collaborate, discuss ideas, or explore exciting opportunities.

ahmadihsan506@gmail.com

LinkedIn

Github