Model Development Phase

Date	15 March 2024
Team ID	738306
Project Title	Employee performance prediction with ML
Maximum Marks	6 Marks

Model Selection Report

Based on the provided metrics for the three models (Linear Regression, Random Forest Regressor, and XGBoost Regressor), we can make the following observations:

1)Linear Regression:

Moderate Mean Squared Error (MSE) values for both training and testing data.

Relatively low R-squared (R2) scores, indicating weaker fit to the data.

Consistent Mean Absolute Error (MAE) values.

2)Random Forest Regressor:

Lowest Mean Squared Error (MSE) on testing data among the three models, indicating better prediction accuracy.

High R-squared (R2) scores on both training and testing data, suggesting a good fit to the data and capturing more variance.

Consistent Mean Absolute Error (MAE) values.

3)XGBoost Regressor:

Moderate Mean Squared Error (MSE) values on both training and testing data.

Lower R-squared (R2) scores compared to Random Forest Regressor, indicating slightly weaker performance in capturing variance.

Consistent Mean Absolute Error (MAE) values.

Conclusion:

Based on the provided metrics, the Random Forest Regressor appears to be the best-performing model. It demonstrates the lowest Mean Squared Error (MSE) on the testing data, indicating superior prediction accuracy. Additionally, it exhibits high R-squared (R2) scores on both training and testing data, suggesting a robust fit to the data and capturing more variance compared to the other models. Therefore, for this specific task, the Random Forest Regressor is recommended for further exploration and deployment.

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Linear regression model	Moderate Mean Squared Error (MSE) values for both training and testing data. Relatively low R-squared (R2) scores, indicating weaker fit to the data. Consistent Mean Absolute Error (MAE) values.	We used every hyper parameter which is used in the data set.	mean squared error in training: 0.021829740434257082 mean squared error in testing: 0.021321517772632737 r2_score in training data: 0.3038198342280549 r2_score in test: 0.1970042499190925 mean_absolute_error in training data : 0.10769706277175743 mean_absolute_error in testing data: 0.10729554202727433
Random forest model	Lowest Mean Squared Error (MSE) on testing data	We used every hyper parameter which is used in the data set.	mean squared error in training: 0.0022752182381708293

	among the three models, indicating better prediction accuracy. High R-squared (R2) scores on both training and testing data, suggesting a good fit to the data and capturing more variance. Consistent Mean Absolute Error (MAE) values.		mean squared error in testing: 0.011925308844873023 r2_score in training: 0.9274401903683915 r2_score in test data: 0.5508775490117057 mean_absolute_error in training data: 0.10769706277175743 mean_absolute_error in testing: 0.10729554202727433
Xgboost	Moderate Mean Squared Error (MSE) values on both training and testing data. Lower R-squared (R2) scores compared to Random Forest Regressor, indicating slightly weaker performance in capturing variance. Consistent Mean Absolute Error (MAE) values.	We used every hyper parameter which is used in the data set.	mean squared error in training: 0.0036951760704128597 mean squared error in testing data: 0.012631486322544742 r2_score in training data 0.8821558003859931 r2_score in test data: 0.5242819980091831 mean_absolute_error in training data: 0.10769706277175743 mean_absolute_error in testing: 0.10729554202727433
