Model Optimization and Tuning Phase Report

Date	31 July 2024
Team ID	740677
Project Title	Software Employee Salary Predictions
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values

Decision Tree	—	—
Random Forest	—	—

Model	Optimized Metric

| Decision Tree | ```
[51] train_r2 = r2_score(y_train, y_train_pred) * 100
     print("Training R²",train_r2)

     # Calculate and print the R² score for the testing data
     test_r2 = r2_score(y_test, y_test_pred) * 100
     print("Testing R²: ",test_r2)

     Training R² 99.88283394123113
     Testing R²:  -306.5997015507768
``` | | |
| KNN | — | — | |
| **XG** Boosting | — | — | |

Performance Metrics Comparison Report (2 Marks):

| Random Forest | ```
y_test_pred=rfr.predict(X_test)
y_train_pred=rfr.predict(X_train)


train_r2 = r2_score(y_train, y_train_pred) * 100
print("Training R²:",train_r2)
#Calculate and print the R² score for the testing data
test_r2 = r2_score(y_test, y_test_pred) * 100
print("Testing R²",test_r2)


Training R²: 86.03987604146623
Testing R² 0.19943667460349257
``` |
| Linear Regression | ```
y_test_pred=reg.predict(X_test)
y_train_pred=reg.predict(X_train)


train_r2 = r2_score(y_train, y_train_pred) * 100
print("Training R²",train_r2)

# Calculate and print the R² score for the testing data
test_r2 = r2_score(y_test, y_test_pred) * 100
print("Testing R²: ",test_r2)

Training R² 1.7368383587085146
Testing R²:  3.8894244665287347
``` |
| XG Boost Regressor | |

```
y_test_pred=xg_reg.predict(X_test)
y_train_pred=xg_reg.predict(X_train)


train_r2 = r2_score(y_train, y_train_pred) * 100
print("Training R²:",train_r2)
#Calculate and print the R² score for the testing data
test_r2 = r2_score(y_test, y_test_pred) * 100
print("Testing R²: ",test_r2)

Training R²: 72.57209200330892
Testing R²:   -99.38207793799099
```

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Decision Tree Model	The Decision tree model is the final model chosen because of its best overall performance compared to other models. It capture the variance in the data very well with minimal prediction error.