# Model Optimization and Tuning Phase Report

| | |
|---|---|
| Date | 25 July 2025 |
| Team ID | NONE |
| Project Title | Employee Performance Prediction using Machine Learning |

## Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase focuses on improving the performance of the Random Forest model, which was identified as the best-performing algorithm in the initial evaluation ($R^2 \approx 0.46$). Hyperparameter tuning was performed using GridSearchCV to find the optimal combination of parameters such as n_estimators, max_depth, min_samples_split, and min_samples_leaf. The tuned model was re-evaluated on the test set to measure improvement in $R^2$ score. This phase ensures that the final model achieves maximum predictive accuracy and generalization, making it suitable for deployment in real-world workforce management scenarios.

## Hyperparameter Tuning Documentation:

| Model | Tuned Hyperparameters | Optimal Values |
|---|---|---|
| Linear Regression | - No specific hyperparameters (default settings)<br>`lr = LinearRegression()`<br>`lr.fit(X_train, y_train)`<br>`lr_pred = lr.predict(X_test)`<br>`lr_r2 = r2_score(y_test, lr_pred)`<br>`print("Linear Regression R² Score:", lr_r2)` | NONE |
| Random Forest | n_estimators: [50, 100, 200]<br>max_depth: [None, 10, 20, 30]<br>min_samples_split: [2,5, 10]<br>min_samples_leaf: [1,2, 4]<br>random_state: 42<br>n_jobs: -1<br>n_estimators: 100<br>max_depth: None<br>min_samples_split: 5<br>min_samples_leaf: 1 | 100,42 |
| XGBoost | n_estimators: [50, 100, 200]<br>learning_rate: [0.01,0.1,0.2]<br>max_depth: [3, 6, 9]<br>subsample: [0.8, 1.0]<br>colsample_bytree: [0.8, 1.0]<br>random_state: 42 | 100,42 |

**Performance Metrics Comparison Report:**

| Model | Optimized Metric |
|---|---|
| Linear Regression | (base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> &<br>Linear Regression R² Score: 0.1681682566306545<br>Random Forest R² Score: 0.44671974539154335<br>XGBoost R² Score: 0.3538597397101051<br>Best model saved to model/best_model.pkl<br>(base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> |
| Random Forest | (base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> &<br>Linear Regression R² Score: 0.1681682566306545<br>Random Forest R² Score: 0.44671974539154335<br>XGBoost R² Score: 0.3538597397101051<br>Best model saved to model/best_model.pkl<br>(base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> |
| XGBoost | (base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> &<br>Linear Regression R² Score: 0.1681682566306545<br>Random Forest R² Score: 0.44671974539154335<br>XGBoost R² Score: 0.3538597397101051<br>Best model saved to model/best_model.pkl<br>(base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> |

**Final Model Selection Justification**

| Final Model | Reasoning |
|---|---|
| Random Forest | The Random Forest model was selected as the final model due to its superior performance during hyperparameter tuning, achieving an $R^2$ score of approximately |