

Model Optimization and Tuning Phase Report

Date	24 June 2025
Team ID	SWUID20250176341
Project Title	Machine Learning Approach for Employee Performance Prediction
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
Linear Regression	<pre># 1. Linear Regression Model model_lr = LinearRegression() model_lr.fit(X_train, y_train) pred_test = model_lr.predict(X_test) mae_lr = mean_absolute_error(y_test, pred_test) mse_lr = mean_squared_error(y_test, pred_test) r2_lr = r2_score(y_test, pred_test) print("\nLinear Regression:") print("MAE:", mae_lr) print("MSE:", mse_lr) print("R² Score:", r2_lr)</pre>	<pre>Linear Regression: MAE: 0.04347336915994117 MSE: 0.004133703766872829 R² Score: 0.8012302992785618</pre>
Random Forest	<pre># 2. Random Forest Model model_rf = RandomForestRegressor(random_state=42) model_rf.fit(X_train, y_train) pred = model_rf.predict(X_test) mae_rf = mean_absolute_error(y_test, pred) mse_rf = mean_squared_error(y_test, pred) r2_rf = r2_score(y_test, pred) print("\nRandom Forest:") print("MAE:", mae_rf) print("MSE:", mse_rf) print("R² Score:", r2_rf)</pre>	<pre>Random Forest: MAE: 0.03253990252410063 MSE: 0.0035580692923302647 R² Score: 0.8289097602855915</pre>

Xgboost Model	<pre># 3. XGBoost Model model_xgb = XGBRegressor(random_state=42, verbosity=0) model_xgb.fit(X_train, y_train) pred3 = model_xgb.predict(X_test) mae_xgb = mean_absolute_error(y_test, pred3) mse_xgb = mean_squared_error(y_test, pred3) r2_xgb = r2_score(y_test, pred3) print("\nXGBoost:") print("MAE:", mae_xgb) print("MSE:", mse_xgb) print("R² Score:", r2_xgb)</pre>	<p>XGBoost:</p> <p>MAE: 0.033858053386211395</p> <p>MSE: 0.0037014407571405172</p> <p>R² Score: 0.8220157623291016</p>
---------------	---	--

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Random Forest	<pre>----- Model Comparison (Based on Test Set) ----- Model MAE MSE R² Score Linear Regression 0.0435 0.0041 0.8012 Random Forest 0.0325 0.0036 0.8289 XGBoost 0.0339 0.0037 0.8220 ✅ Random Forest model and feature order saved successfully. F2 Score: 0.9046454767726161</pre>

Random Forest	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.71</td><td>0.83</td><td>0.77</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.84</td><td>0.73</td><td>0.78</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.78</td><td>169</td></tr><tr><td>macro avg</td><td>0.78</td><td>0.78</td><td>0.77</td><td>169</td></tr><tr><td>weighted avg</td><td>0.78</td><td>0.78</td><td>0.78</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[62, 13], [25, 69]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.71	0.83	0.77	75	Loan will not be Approved	0.84	0.73	0.78	94	accuracy			0.78	169	macro avg	0.78	0.78	0.77	169	weighted avg	0.78	0.78	0.78	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.71	0.83	0.77	75																											
Loan will not be Approved	0.84	0.73	0.78	94																											
accuracy			0.78	169																											
macro avg	0.78	0.78	0.77	169																											
weighted avg	0.78	0.78	0.78	169																											
KNN	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.73</td><td>0.59</td><td>0.65</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.72</td><td>0.83</td><td>0.77</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.72</td><td>169</td></tr><tr><td>macro avg</td><td>0.72</td><td>0.71</td><td>0.71</td><td>169</td></tr><tr><td>weighted avg</td><td>0.72</td><td>0.72</td><td>0.72</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[44, 31], [16, 78]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.73	0.59	0.65	75	Loan will not be Approved	0.72	0.83	0.77	94	accuracy			0.72	169	macro avg	0.72	0.71	0.71	169	weighted avg	0.72	0.72	0.72	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.73	0.59	0.65	75																											
Loan will not be Approved	0.72	0.83	0.77	94																											
accuracy			0.72	169																											
macro avg	0.72	0.71	0.71	169																											
weighted avg	0.72	0.72	0.72	169																											
Gradient Boosting	<pre>print(classification_report(y_test,y_pred))</pre> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>Loan will be Approved</td><td>0.73</td><td>0.85</td><td>0.79</td><td>75</td></tr><tr><td>Loan will not be Approved</td><td>0.86</td><td>0.74</td><td>0.80</td><td>94</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.79</td><td>169</td></tr><tr><td>macro avg</td><td>0.80</td><td>0.80</td><td>0.79</td><td>169</td></tr><tr><td>weighted avg</td><td>0.80</td><td>0.79</td><td>0.79</td><td>169</td></tr></tbody></table> <pre>confusion_matrix(y_test,y_pred)</pre> <pre>array([[64, 11], [24, 70]])</pre>		precision	recall	f1-score	support	Loan will be Approved	0.73	0.85	0.79	75	Loan will not be Approved	0.86	0.74	0.80	94	accuracy			0.79	169	macro avg	0.80	0.80	0.79	169	weighted avg	0.80	0.79	0.79	169
	precision	recall	f1-score	support																											
Loan will be Approved	0.73	0.85	0.79	75																											
Loan will not be Approved	0.86	0.74	0.80	94																											
accuracy			0.79	169																											
macro avg	0.80	0.80	0.79	169																											
weighted avg	0.80	0.79	0.79	169																											

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Gradient Boosting	The Gradient Boosting model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.