

## Model Optimization and Tuning Phase Template

Date	11 July 2024
Team ID	SWTID1720455879
Project Title	Human Resource Management: Predicting Employee Promotions Using Machine Learning
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	Criterion, max_depth, min_samples_split, min_samples_leaf parameters are used in case of Decision Tree for fine-tuning hyperparameters	<p>The values are</p> <pre>dt_params = {     'criterion': ['gini', 'entropy'],     'max_depth': [None, 10, 20, 30],     'min_samples_split': [2, 5, 10],     'min_samples_leaf': [1, 2, 4] }</pre>

Random Forest	n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features parameters are used in case of Random Forest for fine-tuning hyperparameters	The values are <pre>rf_params = {     'n_estimators': [100, 200],     'max_depth': [10, 20],     'min_samples_split': [2, 5],     'min_samples_leaf': [1, 2],     'max_features': ['sqrt', 'log2'] }</pre>
KNN	n_neighbors, weights, metric parameters are used in case of KNN for fine-tuning hyperparameters	The values are <pre>knn_params = {     'n_neighbors': [3, 5, 7],     'weights': ['uniform', 'distance'],     'metric': ['euclidean', 'manhattan'] }</pre>
XGBoost	n_estimators, learning_rate, max_depth, subsample, colsample_bytree, gamma, min_child_weight parameters are used in case of XGBoost for fine- tuning hyperparameters	The values are <pre>xgb_params = {     'n_estimators': [100, 200],       'learning_rate': [0.01, 0.1],     'max_depth': [3, 5],     'subsample': [0.8, 1.0],     'colsample_bytree': [0.8, 1.0],     'gamma': [0, 0.1],     'min_child_weight': [1, 3] }</pre>

### Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric					Optimized Metric					
Decision Tree		precision	recall	f1-score	support		precision	recall	f1-score	support	
	0	0.93	0.93	0.93	15180	0	0.93	0.93	0.93	15180	
	1	0.93	0.93	0.93	14904	1	0.93	0.93	0.93	14904	
	accuracy			0.93	30084	accuracy			0.93	30084	
	macro avg		0.93	0.93	0.93	30084	macro avg		0.93	0.93	30084
	weighted avg		0.93	0.93	0.93	30084	weighted avg		0.93	0.93	30084

Random Forest	precision	recall	f1-score	support	precision	recall	f1-score	support		
	0	0.94	0.96	0.95	15180	0	0.95	0.93	0.94	15180
	1	0.96	0.94	0.95	14904	1	0.93	0.95	0.94	14904
	accuracy			0.95	30084	accuracy			0.94	30084
	macro avg	0.95	0.95	0.95	30084	macro avg	0.94	0.94	0.94	30084
	weighted avg	0.95	0.95	0.95	30084	weighted avg	0.94	0.94	0.94	30084
KNN	precision	recall	f1-score	support	precision	recall	f1-score	support		
	0	0.97	0.82	0.89	15180	0	0.95	0.90	0.93	15180
	1	0.84	0.97	0.90	14904	1	0.91	0.95	0.93	14904
	accuracy			0.89	30084	accuracy			0.93	30084
	macro avg	0.90	0.89	0.89	30084	macro avg	0.93	0.93	0.93	30084
	weighted avg	0.90	0.89	0.89	30084	weighted avg	0.93	0.93	0.93	30084
XGBoost	precision	recall	f1-score	support	precision	recall	f1-score	support		
	0	0.91	0.98	0.94	15180	0	0.90	0.97	0.93	15180
	1	0.98	0.90	0.94	14904	1	0.97	0.89	0.93	14904
	accuracy			0.94	30084	accuracy			0.93	30084
	macro avg	0.94	0.94	0.94	30084	macro avg	0.93	0.93	0.93	30084
	weighted avg	0.94	0.94	0.94	30084	weighted avg	0.93	0.93	0.93	30084

### Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random Forest	The Random Forest model achieved the highest accuracy among all the models tested, indicating that it was the best at correctly predicting employee promotions. Accuracy is a key metric in classification problems as it measures the proportion of correctly predicted instances out of the total instances.