

## Model Optimization and Tuning Phase Template

Date	17 July 2024
Team ID	-
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	<pre> .....# Define the hyperparameters and their possible values .....param_dist = { .....    'criterion': ['gini', 'entropy'], .....    'max_depth': [None, 10, 20, 30, 40, 50], .....    'min_samples_split': randint(2, 11), .....    'min_samples_leaf': randint(1, 5), .....    'max_features': [None, 'auto', 'sqrt', 'log2'] .....} </pre>	<p>Fitting 5 folds for each of 100 candidates, totalling 500 fits</p> <p>Best Parameters: {'criterion': 'gini', 'max_depth': 30, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 4}</p> <p><b>Accuracy: 0.93</b></p>
Random Forest	<pre> .....# Define the hyperparameters and their possible values .....param_dist = { .....    'n_estimators': randint(100, 500), .....    'max_features': ['auto', 'sqrt'], .....    'max_depth': randint(10, 30), .....    'min_samples_split': randint(2, 10), .....    'min_samples_leaf': randint(1, 3), .....    'bootstrap': [True] .....} </pre>	<p>Fitting 3 folds for each of 50 candidates, totalling 150 fits</p> <p>Best Parameters: {'bootstrap': True, 'max_depth': 20, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 400}</p> <p>Confusion Matrix</p> <p><b>Accuracy: 0.95</b></p>

KNN	<pre> ...# Define the hyperparameters and their possible values ...param_dist = { ...    'n_neighbors': randint(1, 30), ...# Number of neighbors ...    'weights': ['uniform', 'distance'], ...# Weight function ...    'metric': ['euclidean', 'manhattan', 'minkowski'] ...# Distance metric ...} ... </pre>	<p>Fitting 3 folds for each of 50 candidates, totalling 150 fits</p> <p>Best Parameters: {'metric': 'manhattan', 'n_neighbors': 4, 'weights': 'distance'}</p> <p><b>Accuracy: 0.91</b></p>
Xgboost	<pre> ...param_dist = { ...    'n_estimators': randint(50, 500), ...# N ...    'learning_rate': uniform(0.01, 0.3), ... ...    'max_depth': randint(3, 15), ...# Maximum ...    'min_child_weight': randint(1, 10), ...# ...    'subsample': uniform(0.5, 0.5), ...# Fra ...    'colsample_bytree': uniform(0.5, 0.5), ...    'gamma': uniform(0, 5) ...# Minimum loss ...} ... </pre>	<p>Fitting 3 folds for each of 50 candidates, totalling 150 fits</p> <p>Best Parameters: {'colsample_bytree': 0.8251598637910, 'gamma': 0.1941737014115, 'learning_rate': 0.0007954402650, 'max_depth': 13, 'min_child_weight': 5, 'n_estimators': 100}</p> <p><b>Accuracy: 0.94</b></p>

## Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric																																																												
Decision Tree	<div>Confusion Matrix: [[13853 1212] [ 878 14141]]</div> <div>Classification Report: <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.94</td><td>0.92</td><td>0.93</td><td>15065</td></tr><tr><td>1</td><td>0.92</td><td>0.94</td><td>0.93</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.93</td><td>30084</td></tr><tr><td>macro avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>30084</td></tr></tbody></table><div>Accuracy: 0.93</div></div>		precision	recall	f1-score	support	0	0.94	0.92	0.93	15065	1	0.92	0.94	0.93	15019	accuracy			0.93	30084	macro avg	0.93	0.93	0.93	30084	weighted avg	0.93	0.93	0.93	30084	<div>Confusion Matrix: [[13966 1099] [ 983 14036]]</div> <div>Classification Report: <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.93</td><td>0.93</td><td>0.93</td><td>15065</td></tr><tr><td>1</td><td>0.93</td><td>0.93</td><td>0.93</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.93</td><td>30084</td></tr><tr><td>macro avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.93</td><td>0.93</td><td>0.93</td><td>30084</td></tr></tbody></table><div>Accuracy: 0.93</div></div>		precision	recall	f1-score	support	0	0.93	0.93	0.93	15065	1	0.93	0.93	0.93	15019	accuracy			0.93	30084	macro avg	0.93	0.93	0.93	30084	weighted avg	0.93	0.93	0.93	30084
	precision	recall	f1-score	support																																																										
0	0.94	0.92	0.93	15065																																																										
1	0.92	0.94	0.93	15019																																																										
accuracy			0.93	30084																																																										
macro avg	0.93	0.93	0.93	30084																																																										
weighted avg	0.93	0.93	0.93	30084																																																										
	precision	recall	f1-score	support																																																										
0	0.93	0.93	0.93	15065																																																										
1	0.93	0.93	0.93	15019																																																										
accuracy			0.93	30084																																																										
macro avg	0.93	0.93	0.93	30084																																																										
weighted avg	0.93	0.93	0.93	30084																																																										
Random Forest	<div>Confusion Matrix: [[14187 878] [ 758 14261]]</div> <div>Classification Report: <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.95</td><td>0.94</td><td>0.95</td><td>15065</td></tr><tr><td>1</td><td>0.94</td><td>0.95</td><td>0.95</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.95</td><td>30084</td></tr><tr><td>macro avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr></tbody></table><div>Accuracy: 0.95</div></div>		precision	recall	f1-score	support	0	0.95	0.94	0.95	15065	1	0.94	0.95	0.95	15019	accuracy			0.95	30084	macro avg	0.95	0.95	0.95	30084	weighted avg	0.95	0.95	0.95	30084	<div>Confusion Matrix: [[14248 817] [ 814 14205]]</div> <div>Classification Report: <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.95</td><td>0.95</td><td>0.95</td><td>15065</td></tr><tr><td>1</td><td>0.95</td><td>0.95</td><td>0.95</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.95</td><td>30084</td></tr><tr><td>macro avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.95</td><td>0.95</td><td>0.95</td><td>30084</td></tr></tbody></table><div>Accuracy: 0.95</div></div>		precision	recall	f1-score	support	0	0.95	0.95	0.95	15065	1	0.95	0.95	0.95	15019	accuracy			0.95	30084	macro avg	0.95	0.95	0.95	30084	weighted avg	0.95	0.95	0.95	30084
	precision	recall	f1-score	support																																																										
0	0.95	0.94	0.95	15065																																																										
1	0.94	0.95	0.95	15019																																																										
accuracy			0.95	30084																																																										
macro avg	0.95	0.95	0.95	30084																																																										
weighted avg	0.95	0.95	0.95	30084																																																										
	precision	recall	f1-score	support																																																										
0	0.95	0.95	0.95	15065																																																										
1	0.95	0.95	0.95	15019																																																										
accuracy			0.95	30084																																																										
macro avg	0.95	0.95	0.95	30084																																																										
weighted avg	0.95	0.95	0.95	30084																																																										

KNN	<p>Confusion Matrix: [[12332 2733] [ 534 14485]]</p> <p>Classification Report:</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.96</td><td>0.82</td><td>0.88</td><td>15065</td></tr><tr><td>1</td><td>0.84</td><td>0.96</td><td>0.90</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.89</td><td>30084</td></tr><tr><td>macro avg</td><td>0.90</td><td>0.89</td><td>0.89</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.90</td><td>0.89</td><td>0.89</td><td>30084</td></tr></tbody></table> <p>Accuracy: 0.89</p>		precision	recall	f1-score	support	0	0.96	0.82	0.88	15065	1	0.84	0.96	0.90	15019	accuracy			0.89	30084	macro avg	0.90	0.89	0.89	30084	weighted avg	0.90	0.89	0.89	30084	<p>Confusion Matrix: [[13156 1909] [ 731 14288]]</p> <p>Classification Report:</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.95</td><td>0.87</td><td>0.91</td><td>15065</td></tr><tr><td>1</td><td>0.88</td><td>0.95</td><td>0.92</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.91</td><td>30084</td></tr><tr><td>macro avg</td><td>0.91</td><td>0.91</td><td>0.91</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.91</td><td>0.91</td><td>0.91</td><td>30084</td></tr></tbody></table> <p>Accuracy: 0.91</p>		precision	recall	f1-score	support	0	0.95	0.87	0.91	15065	1	0.88	0.95	0.92	15019	accuracy			0.91	30084	macro avg	0.91	0.91	0.91	30084	weighted avg	0.91	0.91	0.91	30084
	precision	recall	f1-score	support																																																										
0	0.96	0.82	0.88	15065																																																										
1	0.84	0.96	0.90	15019																																																										
accuracy			0.89	30084																																																										
macro avg	0.90	0.89	0.89	30084																																																										
weighted avg	0.90	0.89	0.89	30084																																																										
	precision	recall	f1-score	support																																																										
0	0.95	0.87	0.91	15065																																																										
1	0.88	0.95	0.92	15019																																																										
accuracy			0.91	30084																																																										
macro avg	0.91	0.91	0.91	30084																																																										
weighted avg	0.91	0.91	0.91	30084																																																										
Xgboost	<p>Confusion Matrix: [[12546 2519] [ 1454 13565]]</p> <p>Classification Report:</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.90</td><td>0.83</td><td>0.86</td><td>15065</td></tr><tr><td>1</td><td>0.84</td><td>0.90</td><td>0.87</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.87</td><td>30084</td></tr><tr><td>macro avg</td><td>0.87</td><td>0.87</td><td>0.87</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.87</td><td>0.87</td><td>0.87</td><td>30084</td></tr></tbody></table> <p>Accuracy: 0.87</p>		precision	recall	f1-score	support	0	0.90	0.83	0.86	15065	1	0.84	0.90	0.87	15019	accuracy			0.87	30084	macro avg	0.87	0.87	0.87	30084	weighted avg	0.87	0.87	0.87	30084	<p>Confusion Matrix: [[14233 832] [ 1067 13952]]</p> <p>Classification Report:</p> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.93</td><td>0.94</td><td>0.94</td><td>15065</td></tr><tr><td>1</td><td>0.94</td><td>0.93</td><td>0.94</td><td>15019</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.94</td><td>30084</td></tr><tr><td>macro avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>30084</td></tr><tr><td>weighted avg</td><td>0.94</td><td>0.94</td><td>0.94</td><td>30084</td></tr></tbody></table> <p>Accuracy: 0.94</p>		precision	recall	f1-score	support	0	0.93	0.94	0.94	15065	1	0.94	0.93	0.94	15019	accuracy			0.94	30084	macro avg	0.94	0.94	0.94	30084	weighted avg	0.94	0.94	0.94	30084
	precision	recall	f1-score	support																																																										
0	0.90	0.83	0.86	15065																																																										
1	0.84	0.90	0.87	15019																																																										
accuracy			0.87	30084																																																										
macro avg	0.87	0.87	0.87	30084																																																										
weighted avg	0.87	0.87	0.87	30084																																																										
	precision	recall	f1-score	support																																																										
0	0.93	0.94	0.94	15065																																																										
1	0.94	0.93	0.94	15019																																																										
accuracy			0.94	30084																																																										
macro avg	0.94	0.94	0.94	30084																																																										
weighted avg	0.94	0.94	0.94	30084																																																										

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

Final Model	Reasoning
Random Forest	I chose the Random Forest model for predicting employee promotions due to its highest accuracy of 95%, outpacing Decision Tree, KNN, and Gradient Boosting. Its robustness, ability to handle overfitting, and insights into feature importance, combined with its capability to manage complex, non-linear data and scale with large datasets, make it a reliable choice. Hyperparameter tuning further enhanced its performance, confirming its effectiveness for this task.