Model Optimization and Tuning Phase Template

Date	17 July 2024
Team ID	xxxxxx
Project Title	Human Resource Management: Predicting Employee Promotions Using Machine Learning
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

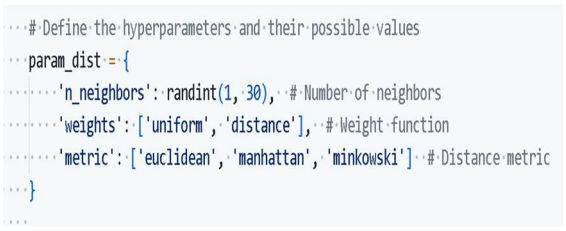
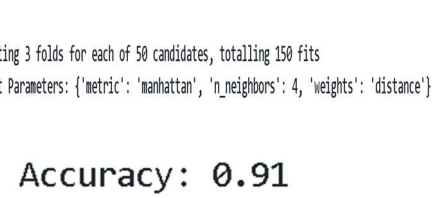
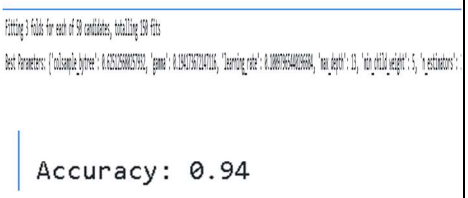
The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (6 Marks):

Model	Tuned Hyperparameters	Optimal Values

Decision Tree	`# 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']`
	`}`
		Fitting 5 folds for each of 100 candidates, totalling 500 fits
		Best Parameters: {'criterion': 'gini', 'max_depth': 30, 'max_features': None, 'min_samples_leaf': 1, 'min_samples_split': 4}
		Accuracy: 0.93

Random Forest	`# 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]`
	`}`
		Fitting 3 folds for each of 50 candidates, totalling 150 fits
		Best Parameters: {'bootstrap': True, 'max_depth': 28, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 419}
		Confusion Matrix:
		Accuracy: 0.95

KNN		
Xgboost		

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Decision Tree	Confusion Matrix: [[13853 1212] [878 14141]] Classification Report: 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 Accuracy: 0.93	Confusion Matrix: [[13966 1099] [983 14036]] Classification Report: 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 Accuracy: 0.93
Random Forest	Confusion Matrix: [[14187 878] [758 14261]] Classification Report: 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 Accuracy: 0.95	Confusion Matrix: [[14248 817] [814 14205]] Classification Report: 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 Accuracy: 0.95

KNN	Confusion Matrix: [[12332 2733] [534 14485]] Classification Report: 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 Accuracy: 0.89	Confusion Matrix: [[13156 1909] [731 14288]] Classification Report: 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 Accuracy: 0.91
Xgboost	Confusion Matrix: [[12546 2519] [1454 13565]] Classification Report: 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 Accuracy: 0.87	Confusion Matrix: [[14233 832] [1067 13952]] Classification Report: 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 Accuracy: 0.94

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.