Model Optimization and Tuning Phase Template

Date	9 July 2024
Team ID	SWTID1720104839
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 hyper parameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyper parameter Tuning Documentation (6 Marks):

Model	Tuned Hyper parameters	Optimal Values
Decision Tree	max_depth, min_samples_split, min_samples_leaf parameters are used in case of Decision Tree for fine-tuning hyper parameters	The values are `param_grid = { 'max_depth': [10, 20, 30, None], 'min_samples_split': [2, 5, 10], 'min_samples_leaf': [1, 2, 4] }`

Random Forest	n_estimators, max_depth,max_leaf_nodes max_features parameters are used in case of Random Forest for fine-tuning hyper parameters	The values are `parameters = { 'n_estimators': [25, 50, 100, 150], 'max_features': ['sqrt', 'log2', None], 'max_depth': [3, 6, 9], 'max_leaf_nodes': [3, 6, 9], }`
KNN	n_neighbors, weights, algorithm parameters are used in case of KNN for fine-tuning hyper parameters	The values are `param_grid = {'n_neighbors': [5,10], 'weights': ['uniform', 'distance'], 'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'], 'p': [1, 2]}`
XG Boost	n_estimators, learning_rate, max_depth, sub sample, parameters are used in case of XG Boost for fine-tuning hyper parameters	The values are `param_grid = { 'max_depth': [3, 5, 7], 'learning_rate': [0.1, 0.01, 0.001], 'subsample': [0.5, 0.7, 1] }`
Gradient Boosting	N_estimators,learning_rate,max_depth parameters are used in case of gradient Boosting for fine-tuning hyper parameters	The values are `param_grid = { 'n_estimators': [50, 100, 200], 'learning_rate': [0.01, 0.1, 0.2], 'max_depth': [3, 5, 7], }`

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Decision Tree	***DecisionTreeClassifier*** confusion matrix [[9187 794] [513 9562]] classification report precision recall f1-score support 0 0.95 0.92 0.93 9981 1 0.92 0.95 0.94 10075 accuracy 0.93 20056 macro avg 0.94 0.93 0.93 20056 weighted avg 0.94 0.93 0.93 20056	***DecisionTreeClassifier_tuned*** confusion matrix [[9300 681] [619 9456]] classification report precision recall f1-score support 0 0.94 0.93 0.93 9981 1 0.93 0.94 0.94 10075 accuracy 0.94 20056 macro avg 0.94 0.94 0.94 20056 weighted avg 0.94 0.94 0.94 20056

Random Forest	***RandomForestClassifier*** confusion matrix [[9432 549] [451 9624]] classification report precision recall f1-score support 0 0.95 0.94 0.95 9981 1 0.95 0.96 0.95 10075 accuracy 0.95 20056 macro avg 0.95 0.95 0.95 20056 weighted avg 0.95 0.95 0.95 20056	***RandomForestClassifier_tuned*** confusion matrix [[6637 3344] [1207 8868]] classification report precision recall f1-score support 0 0.85 0.66 0.74 9981 1 0.73 0.88 0.80 10075 accuracy 0.77 20056 macro avg 0.79 0.77 0.77 20056 weighted avg 0.79 0.77 0.77 20056
KNN	***KNeighborsClassifier*** confusion matrix [[8532 1449] [722 9353]] classification report precision recall f1-score support 0 0.92 0.85 0.89 9981 1 0.87 0.93 0.90 10075 accuracy 0.89 20056 macro avg 0.89 0.89 0.89 20056 weighted avg 0.89 0.89 0.89 20056	***KNeighborsClassifier_tuned*** confusion matrix [[8883 1098] [563 9512]] classification report precision recall f1-score support 0 0.94 0.89 0.91 9981 1 0.90 0.94 0.92 10075 accuracy 0.92 20056 macro avg 0.92 0.92 0.92 20056 weighted avg 0.92 0.92 0.92 20056
XGBoost	***XgBoostingClassifier*** confusion matrix [[9365 616] [1068 9007]] classification report precision recall f1-score support 0 0.90 0.94 0.92 9981 1 0.94 0.89 0.91 10075 accuracy 0.92 20056 macro avg 0.92 0.92 0.92 20056 weighted avg 0.92 0.92 0.92 20056	***XgBoostingClassifier_tuned*** confusion matrix [[9186 795] [1043 9032]] classification report precision recall f1-score support 0 0.90 0.92 0.91 9981 1 0.92 0.90 0.91 10075 accuracy 0.91 20056 macro avg 0.91 0.91 0.91 20056 weighted avg 0.91 0.91 0.91 20056
Gradient Boosting	***GradientBoostingClassifier*** confusion matrix [[8453 1528] [1143 8932]] classification report precision recall f1-score support 0 0.88 0.85 0.86 9981 1 0.85 0.89 0.87 10075 accuracy 0.87 20056 macro avg 0.87 0.87 0.87 20056 weighted avg 0.87 0.87 0.87 20056	***GradientBoostingClassifier_tuned*** confusion matrix [[9502 479] [745 9330]] classification report precision recall f1-score support 0 0.93 0.95 0.94 9981 1 0.95 0.93 0.94 10075 accuracy 0.94 20056 macro avg 0.94 0.94 0.94 20056 weighted avg 0.94 0.94 0.94 20056

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.