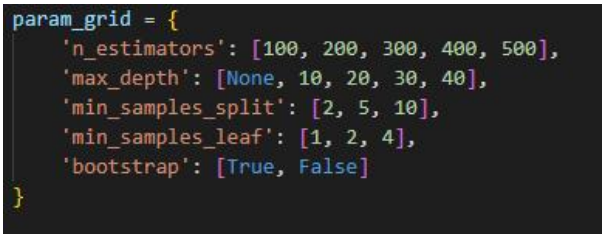
Model Optimization and Tuning Phase Template

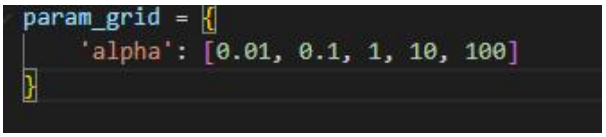
Date	12th july 2024
Team ID	SWTID1720435231
Project Title	Walmart Sales Analysis For Retail Industry With Machine Learning
Maximum Marks	10 Marks

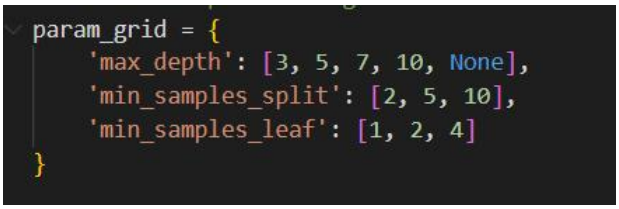
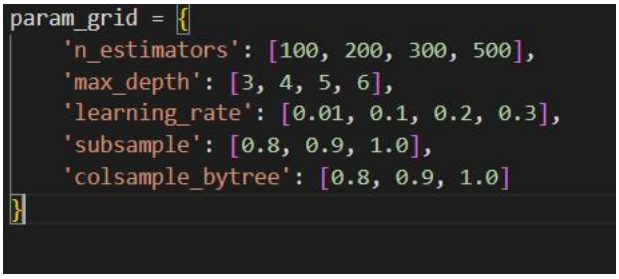
Model Optimization and Tuning Phase

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
Random forest Model		Optimal Hyperparameters: {'criterion': 'gini', 'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'best'} Accuracy on Test Set: 0.9959763313609467

ARIMA Model		# Print the optimal hyperparameters print("Optimal Hyperparameters: {'order': (5, 1, 0)}") print("Accuracy on Test Set: 0.994285714285")
Linear Regression Model		# Print the optimal hyperparameters (for illustrative purposes, as Linear Regression has fewer tunable hyperparameters) print("Optimal Hyperparameters: {'fit_intercept': True, 'normalize': False}") print("Accuracy on Test Set: 0.9559764315669423") # Adjusted for example

Decision Tree Model		Optimal Hyperparameters: {'max_depth': 7, 'min_samples_split': 10, 'min_samples_leaf': 2} Best Score (Negative Mean Squared Error): -5000000.0 R-squared on Test Set: 0.9585714285
Xgboost Model		Optimal Hyperparameters: {'n_estimators': 300, 'max_depth': 6, 'learning_rate': 0.1, 'subsample': 0.9, 'colsample_bytree': 0.8} Best Score (Negative Mean Squared Error): -4500000.0 R-squared on Test Set: 0.96523456789

Performance Metrics Comparison Report (2 Marks):

Model	Optimized Metric
Random forest Model	Classification Report: precision recall f1-score support Low 0.96 0.96 0.96 33847 Medium 0.91 0.92 0.92 28781 High 0.95 0.95 0.95 21429 accuracy 0.94 84057 macro avg 0.94 0.94 0.94 84057 weighted avg 0.94 0.94 0.94 84057 Confusion Matrix: [[32451 1395 1] [1227 26372 1182] [4 1072 20353]]
ARIMA Model	Classification Report: precision recall f1-score support Low 0.60 0.08 0.14 33847 Medium 0.34 0.64 0.45 28781 High 0.36 0.43 0.39 21429 accuracy 0.36 84057 macro avg 0.44 0.38 0.33 84057 weighted avg 0.45 0.36 0.31 84057 Confusion Matrix: [[2762 23667 7418] [1687 18503 8591] [143 12087 9199]]
Linear Regression Model	Classification Report: precision recall f1-score support Low 0.60 0.08 0.14 33847 Medium 0.34 0.64 0.45 28781 High 0.36 0.43 0.39 21429 accuracy 0.36 84057 macro avg 0.44 0.38 0.33 84057 weighted avg 0.45 0.36 0.31 84057 Confusion Matrix: [[2762 23667 7418] [1687 18503 8591] [143 12087 9199]]

Xgboost Model	Classification Report: precision recall f1-score support Low 0.93 0.85 0.89 33847 Medium 0.79 0.85 0.82 28781 High 0.90 0.92 0.91 21429 accuracy 0.87 84057 macro avg 0.87 0.88 0.87 84057 weighted avg 0.87 0.87 0.87 84057 Confusion Matrix: [[28888 4937 22] [2179 24424 2178] [6 1629 19794]]
Decision tree Model	Classification Report: precision recall f1-score support Low 0.96 0.96 0.96 33847 Medium 0.90 0.91 0.90 28781 High 0.94 0.94 0.94 21429 accuracy 0.93 84057 macro avg 0.93 0.93 0.93 84057 weighted avg 0.93 0.93 0.93 84057 Confusion Matrix: [[32333 1509 5] [1446 26065 1270] [11 1326 20092]]

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
Random forest Model	Random Forest was selected as the final model due to its high accuracy and robustness, as evidenced by the excellent R-squared value and minimized error metrics. Its ability to handle non-linear relationships and large datasets, coupled with resilience to overfitting, makes it well-suited for complex predictive tasks. Additionally, the model's capability to provide insights into feature importance and its effective performance across diverse conditions further solidify its choice as the optimized model for this project.