Model Optimization and Tuning Phase Template

Date	07 May 2024
Team ID	722312
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	max_depth, min_samples_split, n_estimators	max_depth=30, min_samples_split=5, n_estimators=150
Decision Tree	random_state	random_state=0
XGBoost	objective, nthread, n_estimators, max_depth, learning_rate	objective='reg:squarederror', nthread=4, n_estimators=500, max_depth=4, learning_rate=0.5

PMDARIMA	train_data, trace, start_p, start_q, start_P, start_Q, max_p, max_q, max_P, max_Q, seasonal, stepwise, error_action, D, max_D	train_data, trace=True, start_p=0, start_q=0, start_P=0, start_Q=0, max_p=10, max_q=10, max_P=10, max_Q=10, seasonal=True, stepwise=False, error_action='ignore', D=1, max_D=10

Performance Metrics Comparison Report (2 Marks):

Model	Baseline Metric	Optimized Metric
Random Forest	Baseline RMSE: 5323.84	Optimized RMSE: 4118.16
Decision Tree	Baseline RMSE: 5323.84	Optimized RMSE: 5323.84
XGBoost	Baseline RMSE: 5533.92	Optimized RMSE: 5533.92
PMDARIMA	Baseline RMSE: 685.54	Optimized RMSE: 685.54

Final Model Selection Justification (2 Marks):

Final Model	Reasoning

Random Forest	The Random Forest model was chosen as the final optimized model due to its superior performance compared to the other models after hyperparameter tuning. It achieved the lowest RMSE of 4118.16 on the test data, indicating better predictive accuracy. The tuned hyperparameters provided a good balance between model complexity and generalization, resulting in improved performance. Additionally, Random Forest models are robust and less prone to overfitting, making them suitable for a variety of datasets, including the Walmart sales analysis project. Hence, the Random Forest model was selected as the final model for this project.