Model Optimization and Tuning Phase Template

|  |  |
| --- | --- |
| Date | 06 July 2024 |
| Team ID | SWTID1720260935 |
| Project Title | ECommerce Shipping Prediction Using Machine Learning |
| Maximum Marks | 10 Marks |

# Model Optimization and Tuning Phase

At this stage, machine learning models are optimized and tuned to achieve optimal performance. To maximize expected accuracy and efficiency, it involves evaluating performance measures, fine-tuning hyperparameters, and supporting the final model selection.

# Hyperparameter Tuning Documentation (6 Marks):

|  |  |  |
| --- | --- | --- |
| **Model** | **Tuned Hyperparameters** | **Optimal Values** |
| SVM | c, kernel, gamma | 1.0, rbf, 0.01 |
| random forest | n\_Estimators, Criterion, Max\_Depth, Max\_features | none,1e-9 |
| KNN | n\_neighbors, weights, algorithm, p | 25, uniform, auto, 2 |
| XGBoost | booster | gbtree |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| ANN | Units, kernel\_initialiser, activation | Input layer: 16, ‘random\_uniform’, ‘relu’  First Hidden Layer:  16, ‘random\_uniform’, ‘relu’  Second Hidden Layer: 8, ‘random\_uniform’, ‘relu’  Output layer: 1, ‘random\_uniform’, ‘relu’ |

**Performance Metrics Comparison Report (2 Marks):**

|  |  |  |
| --- | --- | --- |
| **Model** | **Baseline Metric** | **Optimized Metric** |
| SVM | Accuracy,F1 Score:    : | Accuracy,F1 Score: |

|  |  |  |
| --- | --- | --- |
| random forest | Accuracy,F1 Score: | Accuracy,F1 Score: |
| KNN | Accuracy,F1 Score: | Accuracy,F1 Score: |
| XGBoost | Accuracy,F1 Score: | Accuracy,F1 Score: |

|  |  |  |
| --- | --- | --- |
| ANN | Accuracy,F1 Score: | Accuracy,F1 Score: |

# Final Model Selection Justification (2 Marks):

|  |  |
| --- | --- |
| **Final Model** | **Reasoning** |
| **Random Forest** | Due to its higher performance metrics, the Random Forest model was selected as the most optimized model. It was successful in generating accurate forecasts, as evidenced by its peak accuracy of 68.42%.  It also demonstrated a high precision score of 93.00%, demonstrating its consistency in accurately detecting true positives. The ensemble method used by Random Forest reduces overfitting and enhances generalization to fresh data. Because of these features, Random Forest is the best option for improving delivery time estimates while adhering to the project's goals. |