

Model Optimization and Tuning Phase Template

| | |
|---------------|--|
| Date | 04 July 2024 |
| Team ID | SWTID1720183095 |
| Project Title | ECommerce Shipping Prediction Using Machine Learning |
| Maximum Marks | 10 Marks |

Model Optimization and Tuning Phase

Machine learning models are optimized and tuned during this phase to achieve optimal performance. To improve predicted accuracy and efficiency, it involves fine-tuning hyperparameters, comparing performance metrics, and defending the choice of the final model.

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 | <p>Input layer: 16, 'random_uniform', 'relu'</p> <p>First Hidden Layer:</p> <p>16, 'random_uniform', 'relu'</p> <p>Second Hidden Layer: 8, 'random_uniform', 'relu'</p> <p>Output layer: 1, 'random_uniform', 'relu'</p> |
|-----|---------------------------------------|--|

Performance Metrics Comparison Report (2 Marks):

| Model | Baseline Metric | | | | | Optimized Metric | | | | |
|-------|--------------------|-----------|--------|----------|---------|--------------------|-----------|--------|----------|---------|
| SVM | Accuracy,F1 Score: | | | | | Accuracy,F1 Score: | | | | |
| | | precision | recall | f1-score | support | | precision | recall | f1-score | support |
| | 0 | 0.56 | 0.89 | 0.68 | 895 | 0 | 0.56 | 0.88 | 0.69 | 895 |
| | 1 | 0.87 | 0.51 | 0.65 | 1305 | 1 | 0.86 | 0.53 | 0.66 | 1305 |
| | accuracy | | | 0.67 | 2200 | accuracy | | | 0.67 | 2200 |
| | macro avg | 0.71 | 0.70 | 0.67 | 2200 | macro avg | 0.71 | 0.71 | 0.67 | 2200 |
| | weighted avg | 0.74 | 0.67 | 0.66 | 2200 | weighted avg | 0.74 | 0.67 | 0.67 | 2200 |
| | | | | | | | | | | |

| | | | | | | | | | | |
|---------------|---------------------------------------|-----------|--------|----------|---------|---|-----------|--------|----------|---------|
| random forest | Accuracy,F1 Score: | | | | | Accuracy,F1 Score: | | | | |
| | Baseline Matrix: | | | | | | | | | |
| | | precision | recall | f1-score | support | | precision | recall | f1-score | support |
| | 0 | 0.58 | 0.67 | 0.62 | 895 | 0 | 0.58 | 0.69 | 0.63 | 895 |
| | 1 | 0.75 | 0.66 | 0.70 | 1305 | 1 | 0.75 | 0.66 | 0.70 | 1305 |
| | accuracy | | | 0.67 | 2200 | accuracy | | | 0.67 | 2200 |
| | macro avg | 0.66 | 0.67 | 0.66 | 2200 | macro avg | 0.67 | 0.67 | 0.67 | 2200 |
| | weighted avg | 0.68 | 0.67 | 0.67 | 2200 | weighted avg | 0.68 | 0.67 | 0.67 | 2200 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| KNN | Accuracy,F1 Score: | | | | | Accuracy,F1 Score: | | | | |
| | K-Nearest Neighbors (KNN) Classifier: | | | | | Optimized K-Nearest Neighbors (KNN) Classifier: | | | | |
| | | precision | recall | f1-score | support | | precision | recall | f1-score | support |
| | 0 | 0.56 | 0.63 | 0.60 | 895 | 0 | 0.57 | 0.66 | 0.61 | 895 |
| | 1 | 0.73 | 0.66 | 0.69 | 1305 | 1 | 0.74 | 0.66 | 0.70 | 1305 |
| | accuracy | | | 0.65 | 2200 | accuracy | | | 0.66 | 2200 |
| | macro avg | 0.65 | 0.65 | 0.65 | 2200 | macro avg | 0.66 | 0.66 | 0.66 | 2200 |
| | weighted avg | 0.66 | 0.65 | 0.65 | 2200 | weighted avg | 0.67 | 0.66 | 0.66 | 2200 |
| | | | | | | | | | | |
| | | | | | | | | | | |
| XGBoost | Accuracy,F1 Score: | | | | | Accuracy,F1 Score: | | | | |
| | XGBoost Classifier: | | | | | | | | | |
| | | precision | recall | f1-score | support | | precision | recall | f1-score | support |
| | 0 | 0.57 | 0.91 | 0.70 | 895 | 0 | 0.56 | 0.88 | 0.69 | 895 |
| | 1 | 0.90 | 0.53 | 0.67 | 1305 | 1 | 0.86 | 0.53 | 0.66 | 1305 |
| | accuracy | | | 0.69 | 2200 | accuracy | | | 0.67 | 2200 |
| | macro avg | 0.73 | 0.72 | 0.68 | 2200 | macro avg | 0.71 | 0.71 | 0.67 | 2200 |
| | weighted avg | 0.76 | 0.69 | 0.68 | 2200 | weighted avg | 0.74 | 0.67 | 0.67 | 2200 |
| | | | | | | | | | | |
| | | | | | | | | | | |

| ANN | Accuracy,F1 Score: | | | | | Accuracy,F1 Score: | | | | |
|-----|------------------------|-----------|--------|----------|---------|------------------------|-----------|--------|----------|---------|
| | Classification Report: | | | | | Classification Report: | | | | |
| | | precision | recall | f1-score | support | | precision | recall | f1-score | support |
| | 0 | 0.52 | 0.67 | 0.59 | 1312 | 0 | 0.52 | 0.67 | 0.59 | 1312 |
| | 1 | 0.73 | 0.59 | 0.66 | 1988 | 1 | 0.73 | 0.59 | 0.66 | 1988 |
| | accuracy | | | 0.63 | 3300 | accuracy | | | 0.63 | 3300 |
| | macro avg | 0.63 | 0.63 | 0.62 | 3300 | macro avg | 0.63 | 0.63 | 0.62 | 3300 |
| | weighted avg | 0.65 | 0.63 | 0.63 | 3300 | weighted avg | 0.65 | 0.63 | 0.63 | 3300 |

Final Model Selection Justification (2 Marks):

| Final Model | Reasoning |
|---------------------------|--|
| XGBoost Classifier | <p>The XGBoost Classifier model's better performance metrics led to its selection as the ultimate optimized model. It proved to be effective in creating accurate predictions, as evidenced by its highest accuracy of 69%.</p> <p>It also demonstrated a high precision score of 91.00%, demonstrating its consistency in accurately detecting true positives. The ensemble method used by XGBoost classifier reduces overfitting and enhances generalization to fresh data. Given these features, XGBoost is the best option for improving delivery time forecasts, which is in line with the project's goals.</p> |