

Model Optimization and Tuning Phase Template

| | |
|---------------|---|
| Date | 5 July 2024 |
| Team ID | SWTID1720082658 |
| Project Title | E-commerce Shipping Prediction 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 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 |
|--------------------------|--|--|
| LOGISTIC REGRESSION | Solver,penalty,C,max_iter | Best Cross-Validation Score: 0.6511227264283181 Accuracy with Hyperparameter Tuning and SMOTE: 0.64363636 |
| RANDOM FOREST CLASSIFIER | n_estimators, max_depth, min_samples_split, min_samples_leaf | Best Cross-Validation Score: 0.6839847936339164 Accuracy with Hyperparameter Tuning and SMOTE: 0.6893939393939394 |
| KNN | n_neighbours,weights,metric,p | Best Cross-validation Score: 0.6598739869775356 Accuracy with Hyperparameter Tuning and SMOTE: 0.6633333333333333 |

| | | |
|----------|--|--|
| XG BOOST | n_estimators,max_depth,learning_rate,subsample | <p>Best Cross-Validation Score: 0.682295536990625 Accuracy with Hyperparameter Tuning and SMOTE: 0.698686</p> |
| SVC | Kernel,C,gamma | <p>Best Cross-Validation Score: 0.6680090799389043 Accuracy with Hyperparameter Tuning and SMOTE: 0.670303</p> |

Performance Metrics Comparison Report (2 Marks):

| Model | Baseline Metric | Optimized Metric | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------------------|---|------------------|-----------|---------|----------|---------|---|------|------|------|------|---|------|------|------|------|----------|--|--|------|------|-----------|------|------|------|------|--------------|------|------|------|------|--|--|-----------|--------|----------|---------|---|------|------|------|------|---|------|------|------|------|----------|--|--|------|------|-----------|------|------|------|------|--------------|------|------|------|------|
| LOGISTIC REGRESSION | <div>Accuracy without Hyperparameter Tuning and SMOTE: 0.6284848484848485 Classification Report without Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.56</td><td>0.53</td><td>0.54</td><td>1379</td></tr><tr><td>1</td><td>0.67</td><td>0.70</td><td>0.69</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.63</td><td>3300</td></tr><tr><td>macro avg</td><td>0.62</td><td>0.61</td><td>0.61</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.63</td><td>0.63</td><td>0.63</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix without Hyperparameter Tuning and SMOTE: [[725 654] [572 1349]]</div> | | precision | recall | f1-score | support | 0 | 0.56 | 0.53 | 0.54 | 1379 | 1 | 0.67 | 0.70 | 0.69 | 1921 | accuracy | | | 0.63 | 3300 | macro avg | 0.62 | 0.61 | 0.61 | 3300 | weighted avg | 0.63 | 0.63 | 0.63 | 3300 | <div>Accuracy with Hyperparameter Tuning and SMOTE: 0.64363636 Classification Report with Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.55</td><td>0.76</td><td>0.64</td><td>1379</td></tr><tr><td>1</td><td>0.77</td><td>0.56</td><td>0.65</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.64</td><td>3300</td></tr><tr><td>macro avg</td><td>0.66</td><td>0.66</td><td>0.64</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.68</td><td>0.64</td><td>0.64</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix with Hyperparameter Tuning and SMOTE: [[1053 326] [850 1071]]</div> | | precision | recall | f1-score | support | 0 | 0.55 | 0.76 | 0.64 | 1379 | 1 | 0.77 | 0.56 | 0.65 | 1921 | accuracy | | | 0.64 | 3300 | macro avg | 0.66 | 0.66 | 0.64 | 3300 | weighted avg | 0.68 | 0.64 | 0.64 | 3300 |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.56 | 0.53 | 0.54 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.67 | 0.70 | 0.69 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.63 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.62 | 0.61 | 0.61 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.63 | 0.63 | 0.63 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.55 | 0.76 | 0.64 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.77 | 0.56 | 0.65 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.64 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.66 | 0.66 | 0.64 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.68 | 0.64 | 0.64 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| RANDOM FOREST CLASSIFIER | <div>Accuracy without Hyperparameter Tuning and SMOTE: 0.690000 Classification Report without Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.58</td><td>0.91</td><td>0.71</td><td>1379</td></tr><tr><td>1</td><td>0.89</td><td>0.53</td><td>0.67</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.69</td><td>3300</td></tr><tr><td>macro avg</td><td>0.74</td><td>0.72</td><td>0.69</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.76</td><td>0.69</td><td>0.69</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix without Hyperparameter Tuning and SMOTE: [[1250 129] [894 1027]]</div> | | precision | recall | f1-score | support | 0 | 0.58 | 0.91 | 0.71 | 1379 | 1 | 0.89 | 0.53 | 0.67 | 1921 | accuracy | | | 0.69 | 3300 | macro avg | 0.74 | 0.72 | 0.69 | 3300 | weighted avg | 0.76 | 0.69 | 0.69 | 3300 | <div>Accuracy with Hyperparameter Tuning and SMOTE: 0.6893939393939394 Classification Report with Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.58</td><td>0.95</td><td>0.72</td><td>1379</td></tr><tr><td>1</td><td>0.93</td><td>0.50</td><td>0.65</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.69</td><td>3300</td></tr><tr><td>macro avg</td><td>0.75</td><td>0.73</td><td>0.69</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.78</td><td>0.69</td><td>0.68</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix with Hyperparameter Tuning and SMOTE: [[1306 73] [952 969]]</div> | | precision | recall | f1-score | support | 0 | 0.58 | 0.95 | 0.72 | 1379 | 1 | 0.93 | 0.50 | 0.65 | 1921 | accuracy | | | 0.69 | 3300 | macro avg | 0.75 | 0.73 | 0.69 | 3300 | weighted avg | 0.78 | 0.69 | 0.68 | 3300 |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.58 | 0.91 | 0.71 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.89 | 0.53 | 0.67 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.69 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.74 | 0.72 | 0.69 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.76 | 0.69 | 0.69 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.58 | 0.95 | 0.72 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.93 | 0.50 | 0.65 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.69 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.75 | 0.73 | 0.69 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.78 | 0.69 | 0.68 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| KNN | <div>Accuracy without Hyperparameter Tuning and SMOTE: 0.6696969696969697 Classification Report without Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.58</td><td>0.75</td><td>0.65</td><td>1379</td></tr><tr><td>1</td><td>0.77</td><td>0.61</td><td>0.68</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.67</td><td>3300</td></tr><tr><td>macro avg</td><td>0.68</td><td>0.68</td><td>0.67</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.69</td><td>0.67</td><td>0.67</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix without Hyperparameter Tuning and SMOTE: [[1035 344] [755 1166]]</div> | | precision | recall | f1-score | support | 0 | 0.58 | 0.75 | 0.65 | 1379 | 1 | 0.77 | 0.61 | 0.68 | 1921 | accuracy | | | 0.67 | 3300 | macro avg | 0.68 | 0.68 | 0.67 | 3300 | weighted avg | 0.69 | 0.67 | 0.67 | 3300 | <div>Accuracy with Hyperparameter Tuning and SMOTE: 0.6633333333333333 Classification Report with Hyperparameter Tuning and SMOTE:</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.57</td><td>0.81</td><td>0.67</td><td>1379</td></tr><tr><td>1</td><td>0.80</td><td>0.56</td><td>0.66</td><td>1921</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.66</td><td>3300</td></tr><tr><td>macro avg</td><td>0.68</td><td>0.68</td><td>0.66</td><td>3300</td></tr><tr><td>weighted avg</td><td>0.70</td><td>0.66</td><td>0.66</td><td>3300</td></tr></tbody></table> <div>Confusion Matrix with Hyperparameter Tuning and SMOTE: [[1111 268] [843 1078]]</div> | | precision | recall | f1-score | support | 0 | 0.57 | 0.81 | 0.67 | 1379 | 1 | 0.80 | 0.56 | 0.66 | 1921 | accuracy | | | 0.66 | 3300 | macro avg | 0.68 | 0.68 | 0.66 | 3300 | weighted avg | 0.70 | 0.66 | 0.66 | 3300 |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.58 | 0.75 | 0.65 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.77 | 0.61 | 0.68 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.67 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.68 | 0.68 | 0.67 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.69 | 0.67 | 0.67 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | precision | recall | f1-score | support | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0 | 0.57 | 0.81 | 0.67 | 1379 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 0.80 | 0.56 | 0.66 | 1921 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| accuracy | | | 0.66 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| macro avg | 0.68 | 0.68 | 0.66 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| weighted avg | 0.70 | 0.66 | 0.66 | 3300 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | |
|----------|--|--|
| XG BOOST | <pre> Accuracy without Hyperparameter Tuning and SMOTE: 0.644545 Classification Report without Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.57 0.62 0.59 1379 1 0.71 0.66 0.68 1921 accuracy 0.64 0.64 0.64 3300 macro avg 0.64 0.64 0.64 3300 weighted avg 0.65 0.64 0.65 3300 Confusion Matrix without Hyperparameter Tuning and SMOTE: [[853 526] [647 1274]] </pre> | <pre> Accuracy with Hyperparameter Tuning and SMOTE: 0.690606 Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.58 0.96 0.72 1379 1 0.94 0.50 0.65 1921 accuracy 0.69 0.69 0.69 3300 macro avg 0.76 0.73 0.69 3300 weighted avg 0.79 0.69 0.68 3300 Confusion Matrix with Hyperparameter Tuning and SMOTE: [[1320 59] [962 959]] </pre> |
| SVC | <pre> Accuracy without Hyperparameter Tuning and SMOTE: 0.668788 Classification Report without Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.57 0.82 0.67 1379 1 0.81 0.56 0.66 1921 accuracy 0.67 0.67 0.67 3300 macro avg 0.69 0.69 0.67 3300 weighted avg 0.71 0.67 0.67 3300 Confusion Matrix without Hyperparameter Tuning and SMOTE: [[1129 250] [843 1078]] </pre> | <pre> Accuracy with Hyperparameter Tuning and SMOTE: 0.670303 Classification Report with Hyperparameter Tuning and SMOTE: precision recall f1-score support 0 0.56 0.92 0.70 1379 1 0.90 0.49 0.63 1921 accuracy 0.67 0.67 0.67 3300 macro avg 0.73 0.71 0.67 3300 weighted avg 0.76 0.67 0.66 3300 Confusion Matrix with Hyperparameter Tuning and SMOTE: [[1274 105] [983 938]] </pre> |

Final Model Selection Justification (2 Marks):

| Final Model | Reasoning |
|--------------------------|--|
| RANDOM FOREST CLASSIFIER | The RANDOM FOREST CLASSIFIER model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model |