**For Dataset2:**

**Table 1 ( For Normal Split [ 80:20, 90:10, 85:15 ] )**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Split Ratio** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random forest | Train 90%, Test 10% | 98.50 | 0 | - | - |
| Random Forest | Train 80%, Test 20% | 94.77 | 50 | 57.14 | 53.33 |
| Random Forest | Train 75%, Test 25% | 94.61 | 53.84 | 70 | 60.86 |
| Random Forest | Train 85%, Test 15% | 97.02 | 50 | 33.33 | 40 |
| Logistic Regression | Train 90%, Test 10% | 98.50 | 0 | - | - |
| Logistic Regression | Train 80%, Test 20% | 97.76 | 75 | 85.71 | 79.99 |
| Logistic Regression | Train 75%, Test 25% | 93.41 | 46.66 | 70 | 56 |
| Logistic Regression | Train 85%, Test 15% | 95.04 | 25 | 33.33 | 28.57 |
| SVM(Linear) | Train 90%, Test 10% | 98.50 | 0 | - | - |
| SVM(Linear) | Train 80%, Test 20% | 96.26 | 60 | 85.71 | 70.58 |
| SVM(Linear) | Train 75%, Test 25% | 92.81 | 44.44 | 80 | 57.14 |
| SVM(Linear) | Train 85%, Test 15% | 96.03 | 42.85 | 100 | 60 |
| Decision tree | Train 90%, Test 10% | 97.01 | 0 | - | - |
| Decision tree | Train 80%, Test 20% | 95.52 | 54.54 | 85.71 | 66.66 |
| Decision tree | Train 75%, Test 25% | 93.41 | 47.05 | 80 | 59.25 |
| Decision tree | Train 85%, Test 15% | 95.04 | 33.33 | 66.66 | 44.44 |

**Observation**: Take the best split ratio for each algorithm basis on the classification metrics and do the following tables with that.

Updated Algorithm1: Random Forest

Best Split=85:15

Reason: Random forest with 85:15 split works better than other split set.

Updated Algorithm2: Logistic Regression

Best Split=80:20

Reason: Logistic Regression with 80:20 works far better than other split set.

Updated Algorithm3: SVM(Linear)

Best Split= 85:15

Reason: SVM(Linear ) with 85:15 works better than other split set

Updated Algorithm4: Decision Tree

Best Split=85:15

Reason: Here we can see Decision tree works better with 85:15splits other than other other split set.

**Table 2 (For Model Optimization using Hyperparameter Tuning) (With out Feature sealection)**

**[CV = best CV techniques for each algo)]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Model Optimization** | **Accuracy** | **Precision** | **F1-Score** | **Recall** |
| Decision tree | GridSearchCV | 99 | 100 | 80 | 66.66 |
| Decision tree | RandomizedSearchCV | 97.02 | 50 | 76.33 | 33.33 |
| Decision tree | Bayesian Optimization | 99.00 | 75 | 85.74 | 100 |
| Decision tree | Gradient-based Optimization | 98.01 | 87.32 | 50 | 33.33 |
| Random Forest | GridSearchCV | 99 | 100 | 80 | 66.66 |
| Random Forest | RandomizedSearchCV | 99 | 100 | 80 | 66.66 |
| Random Forest | Bayesian Optimization | 98.01 | 87.32 | 50 | 33.33 |
| Random Forest | Gradient-based Optimization | 99 | 100 | 80 | 66.66 |
| Logistic Regression | GridSearchCV | 98.01 | 87.32 | 50 | 33.33 |
| Logistic Regression | RandomizedSearchCV | 99.00 | 75 | 85.74 | 100 |
| Logistic Regression | Bayesian Optimization | 97.02 | 50 | 76.33 | 33.33 |
| Logistic Regression | Gradient-based Optimization | 99.00 | 75 | 85.74 | 100 |
| SVM (Linear) | GridSearchCV | 98.01 | 87.32 | 50 | 33.33 |
| SVM (Linear) | RandomizedSearchCV | 97.02 | 50 | 76.33 | 33.33 |
| SVM (Linear) | Bayesian Optimization | 99.00 | 75 | 85.74 | 100 |
| SVM (Linear) | Gradient-based Optimization | 98.01 | 87.32 | 50 | 33.33 |

**Observation**: Take the best Model Optimization using Hyperparameter tuningtechnique for each algorithm basis on the classification metrics and do the following table with that.

**Updated Algorithm1: Random Forest**

Best Split= 85:15

Best CV=Gradient-based Optimization, GridSearchCV and RandomizedSearch CV

**Updated Algorithm2: Logistic Regression**

Best Split= 85:15

Best CV= Bayesian Optimization and RandomizedSearch CV

**Updated Algorithm3: SVM (Linear)**

Best Split= 85:15

Best CV= Bayesian Optimization

**Updated Algorithm4: Decision tree**

Best Split= 85:15

Best CV= Bayesian Optimization and GridSearch CV

**Table 3 (Apply feature selection Method)**

**[After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms]**

**(Split – 85:15)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Random Forest | 95.04 | 50 | 40 | 44.44 |
| Logistic Regression | 96.03 | 60 | 60 | 60 |
| SVM (Linear) | 97.02 | 62.50 | 100 | 76.92 |
| Decision tree | 97.02 | 66.66 | 80 | 72.72 |

**Observation**: Take the best feature selection technique for each algorithm basis on the classification metrics and do the following tables with that.

**Updated Algorithm1: Random Forest**

Best Split= 85:15

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm2: Logistic Regression**

Best Split= 85:15

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm3: SVM (Linear)**

Best Split= 85:15

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm4: Decision Tree**

Best Split= 85:15

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Table 4 ( Perform Model Optimization using Hyperparameter After Feature Selection)**

**[CV = best CV techniques for each algo)]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Model Optimization** | **Accuracy** | **Precision** | **F1-Score** | **Recall** |
| Decision tree | GridSearchCV | 98.01 | 71.42 | 83.33 | 100 |
| Decision tree | RandomizedSearchCV | 91.08 | 16.66 | 18.18 | 20 |
| Decision tree | Bayesian Optimization | 98.01 | 71.42 | 83.33 | 100 |
| Decision tree | Gradient-based Optimization | 97.02 | 75 | 66.66 | 60 |
| Random Forest | GridSearchCV | 95.04 | 50 | 44.44 | 40 |
| Random Forest | RandomizedSearchCV | 95.04 | 50 | 44.44 | 40 |
| Random Forest | Bayesian Optimization | 98.01 | 71.42 | 83.33 | 100 |
| Random Forest | Gradient-based Optimization | 95.04 | 50 | 44.44 | 40 |
| Logistic Regression | GridSearchCV | 96.03 | 60 | 60 | 60 |
| Logistic Regression | RandomizedSearchCV | 96.03 | 57.14 | 66.66 | 80 |
| Logistic Regression | Bayesian Optimization | 97.02 | 100 | 57.14 | 40 |
| Logistic Regression | Gradient-based Optimization | 97.02 | 66.66 | 72.72 | 80 |
| SVM (Linear) | GridSearchCV | 96.03 | 60 | 60 | 60 |
| SVM (Linear) | RandomizedSearchCV | 97.02 | 62.50 | 76.92 | 100 |
| SVM (Linear) | Bayesian Optimization | 97.02 | 62.50 | 76.92 | 100 |
| SVM (Linear) | Gradient-based Optimization | 97.02 | 62.50 | 76.92 | 100 |

**Observation**: Take the best Model Optimization using Hyperparameter tuningtechnique for each algorithm basis on the classification metrics and do the following table with that.

**Updated Algorithm1: Random Forest**

Best Split= 85:15

Best CV= Bayesian Optimization

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm2: Logistic Regression**

Best Split= 85:15

Best CV= Bayesian Optimization and Gradient-based Optimization

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm3: SVM (Linear)**

Best Split= 85:15

Best CV=RandomizedSearchCV, Bayesian Optimization and Gradient-based Optimization

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**Updated Algorithm4: Decision Tree**

Best Split= 85:15

Best CV= Grid Search CV and Bayesian Optimization

Best Feature selection= After running 5 Feature Selection Method i.e., K-Best, Mutual Info Classif, Chi-square, Correlation matrix we have inferred by Selecting 5 Best Features from the Dataset and then run the previous algorithms

**# Conclusion Model with feature selection VS Model without Feature Selection**

**# model performance with hyperparameter tuning VS model performance without hyperparameter tuing**

**Table 5 (For Choosing best model)**

**Reason:**

|  |  |
| --- | --- |
| **Best algorithm Name** |  |
| **Model description** |  |
|  | |
| **Classification metric 1** |  |
| **Classification metric 2** |  |
| **Classification metric 3** |  |
| **Classification metric 4** |  |
| **Classification metric 5** |  |
| **Classification metric 6** |  |
| **Classification metric 7** |  |
| **Classification metric 8** |  |
| **Classification metric 9** |  |
| **Classification metric 10** |  |
| **Classification metric 11** |  |
| **Classification metric 12** |  |
| **Classification metric 13** |  |
| **Classification metric 14** |  |
| **Classification metric 15** |  |
| **Classification metric 16** |  |
| **Classification metric 17** |  |
| **Classification metric 18** |  |
| **Classification metric 19** |  |
| **Classification metric 20** |  |
| **Classification metric 21** |  |
| **Classification metric 22** |  |