For Feature Concentration

Raw DataSet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Recall | F1 | Auc | Mcc |
| Decision Tree （CART） | 0.5623 | 0.2905 | 0.2734 | 0.4848 | 0.2064 |
| Random Forest | 0.6800 | 0.2595 | 0.3007 | 0.7086 | 0.2085 |
| LogisticRegression\_liblinear | 0.6750 | 0.3665 | 0.3895 | 0.5319 | 0.1500 |
| LogisticRegression\_lbfgs | 0.5795 | 0.2643 | 0.2279 | 0.4236 | 0.0291 |
| LogisticRegression\_newton-cg | 0.6656 | 0.1476 | 0.2141 | 0.5746 | 0.1887 |
| LogisticRegression\_sag | 0.5180 | 0.3929 | 0.3175 | 0.4716 | 0.0658 |
| LogisticRegression\_saga | 0.5093 | 0.4976 | 0.3606 | 0.5064 | 0.1284 |
| SVM\_SVC\_linear | 0.6712 | 0.3833 | 0.4039 | 0.5289 | 0.2262 |
| SVM\_SVC\_poly | 0.6844 | 0.3261 | 0.3509 | 0.5153 | 0.1959 |
| Naïve\_bayes\_GaussianNB | 0.6208 | 0.3619 | 0.3197 | 0.5914 | 0.1125 |
| Naïve\_bayes\_BernoulliNB | 0.6239 | 0.6595 | 0.5040 | 0.7219 | 0.3483 |
| XGBoost | 0.6633 | 0.4190 | 0.4061 | 0.7080 | 0.2314 |
| AdaBoost | 0.6755 | 0.4309 | 0.4224 | 0.7271 | 0.3630 |
| KNN | 0.6678 | 0.3286 | 0.3722 | 0.6755 | 0.3205 |

PCA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Recall | F1 | Auc | Mcc |
| Decision Tree （CART） | 0.7113 | 0.5548 | 0.5243 | 0.6672 | 0.5078 |
| Random Forest | 0.7134 | 0.4167 | 0.4276 | 0.7372 | 0.2402 |
| LogisticRegression\_liblinear | 0.7033 | 0.1048 | 0.1563 | 0.6547 | 0.1628 |
| LogisticRegression\_lbfgs | 0.7032 | 0.1048 | 0.1563 | 0.6547 | 0.2286 |
| LogisticRegression\_newton-cg | 0.7032 | 0.1048 | 0.1563 | 0.6547 | 0.2123 |
| LogisticRegression\_sag | 0.7032 | 0.1048 | 0.1563 | 0.6547 | 0.1628 |
| LogisticRegression\_saga | 0.7032 | 0.1048 | 0.1563 | 0.6547 | 0.1563 |
| SVM\_SVC\_linear | 0.6985 | 0.3045 | 0.3785 | 0.6512 | 0.2366 |
| SVM\_SVC\_poly | 0.6985 | 0.3126 | 0.3385 | 0.6341 | 0.2455 |
| Naïve\_bayes\_GaussianNB | 0.7115 | 0.4929 | 0.4669 | 0.6810 | 0.4665 |
| Naïve\_bayes\_BernoulliNB | 0.7436 | 0.4190 | 0.4652 | 0.6871 | 0.3874 |
| XGBoost | 0.7422 | 0.3976 | 0.4565 | 0.7813 | 0.3917 |
| AdaBoost | 0.6805 | 0.4847 | 0.4297 | 0.6565 | 0.1669 |
| KNN | 0.6835 | 0.3414 | 0.3896 | 0.6446 | 0.1659 |

F16

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Recall | F1 | Auc | Mcc |
| Decision Tree （CART） | 0.7467 | 0.7333 | 0.5665 | 0.7619 | 0.4071 |
| Random Forest | 0.7514 | 0.5238 | 0.5557 | 0.7924 | 0.5855 |
| LogisticRegression\_liblinear | 0.7209 | 0.2786 | 0.3235 | 0.7412 | 0.2923 |
| LogisticRegression\_lbfgs | 0.7384 | 0.5500 | 0.4364 | 0.7519 | 0.2432 |
| LogisticRegression\_newton-cg | 0.7249 | 0.2885 | 0.3212 | 0.7485 | 0.2923 |
| LogisticRegression\_sag | 0.7158 | 0.5881 | 0.4151 | 0.6901 | 0.2343 |
| LogisticRegression\_saga | 0.7158 | 0.5881 | 0.4151 | 0.6901 | 0.2343 |
| SVM\_SVC\_linear | 0.7515 | 0.4095 | 0.4778 | 0.7372 | 0.2976 |
| SVM\_SVC\_poly | 0.7643 | 0.3904 | 0.3647 | 0.7222 | 0.2631 |
| Naïve\_bayes\_GaussianNB | 0.7152 | 0.8024 | 0.5222 | 0.8076 | 0.5110 |
| Naïve\_bayes\_BernoulliNB | 0.7204 | 0.9571 | 0.6315 | 0.8038 | 0.4442 |
| XGBoost | 0.7581 | 0.4881 | 0.4887 | 0.7829 | 0.5490 |
| AdaBoost | 0.7165 | 0.5334 | 0.4945 | 0.7773 | 0.3729 |
| KNN | 0.7034 | 0.3785 | 0.4092 | 0.7227 | 0.3977 |