|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reported information | Test type | Decision Tree | Logistic Regression | Naïve Bayes | K-NN |  |
| Accuracy with default parameters | Entire dataset | 67.05 | 71.06 | 51.13 | 99.97 |  |
| 50% | 66.20% | 71.57 | 52.29 | 100 |  |
| Cross-validation | 66.55% +/- 0.38% | 70.53% +/- 1.55% | 50.96% +/- 2.01% | 62.67% +/- 2.34% |  |
| Accuracy with best parameters | Entire dataset | 67.58 | 72.04 | 51.13 | 99.97 |  |
| 50% | 66.7 | 71.97 | 52.29 | 100 |  |
| Cross-validation | 66.66% +/- 0.29% | 70.81% +/- 2.80% | 50.99% +/- 3.61% | 66.43% +/- 2.53% |  |
| List names of parameters | | Maximal Depth  Apply Pruning  Confidence  Apply Pre-pruning  Minimal Gain  Minimal leaf size  Minimal Size for split  Number of Pre-pruning | Reproducible  Use Regularization  Standardize  Non-neg Coefficients  Add intercept  Compute p-values  Remove collinear columns  Missing values handling  Max iterations  Max runtime Second | Laplace Correction | K  Weighted Vote  Measure Type  Numerical measure |  |
| List selected best values of parameters (in the same order as in the list of names) | Entire dataset | 25  On  0.25  Off | On  Max Number: 2  Off  Off  Off  Off  Off  Off  Skip  10 | On | 1  On  Numerical Measure  Euclidean Distance |  |
| 50% | 25  On  0.3  On  0.1  3  3  2 | On  Max Number: 6  Off  On  On  On  On  On  Mean Imputation  0 | On | 1  On  Numerical Measure  Euclidean Distance |  |
| Cross-validation | 20  On  0.2  On  0.142  2  4  2 | Split on batch: Off  Leave one out: Off  Number of folds: 15  Sampling Type: Stratified Sampling  Use Local Random seed: Off  Enable parallel execution: On | Number of Folds: 15  Sampling type: Stratified Sampling | Number of Folds:11  Sampling Type: Linear Sampling  1  On  Numerical Measure  MaxProductSimilarity |  |