Table 1: Description of Dataset

|  |  |  |
| --- | --- | --- |
|  | Number of Attributes | Number of Instances |
| Diabetes Symptom Dataset | 18 | 435 |

|  |  |
| --- | --- |
| Attributes | Values |
| Age |  |
| Gender | 1. Male 2. Female |
| SBP |  |
| DBP |  |
| Diabetes | 1. No, 2. Yes |
| Ischemic Heart Disease | 1. No, 2. Yes |
| Family History of stroke | 1. No, 2. Yes |
| Alcoholic | 1. No, 2. Yes |
| Less Physically Active | 1. No, 2. Yes |
| Smoking | 1. No, 2. Yes |
| Stress and depression | 1. No, 2. Yes |
| TG↑ | 1. No, 2. Yes |
| LDL↑ | 1. No, 2. Yes |
| HDL↓ | 1. No, 2. Yes |
| Situated Fat↑ | 1. No, 2. Yes |
| Fibre↓ | 1. No, 2. Yes |
| CKD-Kidney | 1. No, 2. Yes |
| Class Attribute | 1. Stroke, 2. Non-stroke |

Table 2: Description of Attribute

Table 3: Comparison of evaluation metrics using 10 fold cross validation and percentage split (80:20)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Evaluation Metrics | 10- Fold Cross Validation | | | | | | | Percentage Split (80:20) | | | | | | |
| NB | SVM | LG | J48 | RT | RF | ANN | NB | SVM | LG | J48 | RT | RF | ANN |
| Total number of the instance | 435 | 435 | 435 | 435 | 435 | 435 | 435 | 87 | 87 | 87 | 87 | 87 | 87 | 87 |
| Correctly Classified  Instance | 348 | 342 | 351 | 355 | 341 | 361 | 356 | 70 | 67 | 68 | 68 | 67 | 66 | 68 |
| 80% | 78.62% | 80.69% | 81.61% | 78.39% | 82.99% | 81.84% | 80.46% | 77.01% | 78.16% | 78.16% | 77.01% | 75.86% | 78.16% |
| Incorrectly classified instance | 87 | 93 | 84 | 80 | 94 | 74 | 79 | 17 | 20 | 19 | 19 | 20 | 21 | 19 |
| 20% | 21.38% | 19.31% | 18.39% | 21.61% | 17.01% | 18.16% | 19.54% | 22.99% | 21.84% | 21.84% | 22.99% | 24.13% | 21.84% |

Table 4: Comparison of performance parameters using 10- Fold Cross-Validation

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Performance Parameters | Class | Weighted Average | | | | | | | |
| NB | SVM | LG | J48 | RT | RF | ANN |
| TP Rate | Positive | 0.874 | 1.000 | 0.936 | 0.909 | 0.845 | 0.909 | 0.898 |
| Negative | 0.527 | 0.000 | 0.333 | 0.473 | 0.559 | 0.538 | 0.527 |
| Weighted Average | 0.800 | 0.786 | 0.807 | 0.816 | 0.784 | 0.830 | 0.818 |
| FP Rate | Positive | 0.473 | 1.000 | 0.667 | 0.527 | 0.441 | 0.462 | 0.473 |
| Negative | 0.126 | 0.000 | 0.064 | 0.091 | 0.155 | 0.091 | 0.102 |
| Weighted Average | 0.399 | 0.786 | 0.538 | 0.434 | 0.380 | 0.383 | 0.394 |
| Precision | Positive | 0.872 | 1.000 | 0.838 | 0.864 | 0.876 | 0.879 | 0.875 |
| Negative | 0.533 | 0.000 | 0.585 | 0.587 | 0.495 | 0.617 | 0.583 |
| Weighted Average | 0.799 | 0.786 | 0.784 | 0.805 | 0.794 | 0.823 | 0.812 |
| Recall | Positive | 0.874 | 1.000 | 0.936 | 0.909 | 0.845 | 0.909 | 0.898 |
| Negative | 0.527 | 0.000 | 0.333 | 0.473 | 0.559 | 0.538 | 0.527 |
| Weighted Average | 0.800 | 0.786 | 0.807 | 0.816 | 0.784 | 0.830 | 0.818 |

Table 5: Comparison of performance parameters using Percentage Split (80:20)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Performance Parameters | Class | Weighted Average | | | | | | | |
| NB | SVM | LG | J48 | RT | RF | ANN |
| TP Rate | Positive | 0.864 | 1.000 | 0.939 | 0.924 | 0.864 | 0.894 | 0.939 |
| Negative | 0.619 | 0.048 | 0.286 | 0.333 | 0.476 | 0.333 | 0.286 |
| Weighted Average | 0.805 | 0.770 | 0.782 | 0.782 | 0.770 | 0.759 | 0.782 |
| FP Rate | Positive | 0.381 | 0.952 | 0.714 | 0.667 | 0.524 | 0.667 | 0.714 |
| Negative | 0.136 | 0.000 | 0.061 | 0.076 | 0.136 | 0.106 | 0.061 |
| Weighted Average | 0.322 | 0.722 | 0.557 | 0.524 | 0.430 | 0.531 | 0.557 |
| Precision | Positive | 0.877 | 0.767 | 0.805 | 0.813 | 0.838 | 0.808 | 0.805 |
| Negative | 0.591 | 1.000 | 0.600 | 0.583 | 0.526 | 0.500 | 0.600 |
| Weighted Average | 0.808 | 0.824 | 0.756 | 0.758 | 0.763 | 0.734 | 0.756 |
| Recall | Positive | 0.864 | 1.000 | 0.939 | 0.924 | 0.864 | 0.894 | 0.939 |
| Negative | 0.619 | 0.048 | 0.286 | 0.333 | 0.476 | 0.333 | 0.286 |
| Weighted Average | 0.805 | 0.770 | 0.782 | 0.782 | 0.770 | 0.759 | 0.782 |

Fig. 1: Performance of Classification Algorithms Using 10- Fold Cross-Validation Technique

Fig. 2: Performance of Classification Algorithms Using Percentage Split Technique