Michael Cummings

11/30/2022  
CS-453 Data Mining

WEKA Tables

Classifying accuracy based on attribute “is\_rising”:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm/ Test Option | OneR | ZeroR | PART | J48 | REPTree | RandomTree | Naïve Bayes |
| Cross-Validation: 5 | 86.62%\* | 86.62%\* | 99.3% | 99.3% | 86.62%\* | 94.37% | 95.07% |
| Cross-Validation: 10 | 86.62%\* | 86.62%\* | 99.3% | 99.3% | 86.62%\* | 92.96% | 93.66% |
| Cross-Validation: 20 | 86.62%\* | 86.62%\* | 99.3% | 99.3% | 86.62%\* | 91.55% | 93.66% |
| Percent Split: 50% | 87.32%\* | 87.32%\* | 97.2% | 97.2% | 87.32%\* | 87.32% | 92.96% |
| Percent Split: 66% | 87.5%\* | 87.5%\* | 100% | 100% | 87.5%\* | 100% | 93.75% |
| Percent Split: 75% | 82.85%\* | 82.86%\* | 100% | 100% | 82.86%\* | 85.71% | 88.57% |

\* All “False” instances classified incorrectly

Classifying based on attribute “state”:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm/ Test Option | OneR | ZeroR | PART | J48 | REPTree | RandomTree | Naïve Bayes |
| Cross-Validation: 5 | 3.52% | 11.27% | 66.9% | 69.01% | 65.49% | 63.38% | 64.08% |
| Cross-Validation: 10 | 3.52% | 11.27% | 64.79% | 66.9% | 64.79% | 59.86% | 64.79% |
| Cross-Validation: 20 | 3.52% | 11.27% | 63.38% | 66.9% | 64.79% | 59.86% | 64.79% |
| Percent Split: 50% | 2.82% | 9.86% | 36.62% | 40.84% | 32.39% | 42.25% | 42.25% |
| Percent Split: 66% | 4.17% | 14.58% | 43.75% | 58.33% | 54.17% | 39.58% | 52.08% |
| Percent Split: 75% | 2.86% | 11.43% | 71.43% | 65.71% | 40% | 80% | 57.14% |

Clustering instances based on “is\_rising” class:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algorithm | Cobweb | EM | SimpleKMeans = 3 | SimpleKMeans = 4 | SimpleKMeans = 6 |
| Accuracy | 2.11% | 35.2% | 62.67% | 43.66% | 40.85% |
| No. Clusters | 198 | 8 | 3 | 4 | 6 |
| Hierarchal/Flat? | Hierarchal | Flat | Flat | Flat | Flat |
| Deterministic/Probabilistic? | Deterministic | Deterministic | Deterministic | Deterministic | Deterministic |
| Disjoint/Overlapping? | Overlapping | Disjoint | Disjoint | Disjoint | Disjoint |

Clustering instances based on “state” class:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algorithm | Cobweb | EM | SimpleKMeans = 3 | SimpleKMeans = 6 | SimpleKMeans = 10 |
| Accuracy | [Error] | 24.65% | 13.38% | 26.06% | 29.58% |
| No. Clusters | 192 | 5 | 3 | 6 | 10 |

Numerical prediction based on “msl\_trend\_mm\_year” class:

|  |  |
| --- | --- |
|  | Linear Regression |
| Cross-Validation: 5 | 58% |
| Cross-Validation: 10 | 59% |
| Cross-Validation: 20 | 60% |
| Percent Split: 50% | 51% |
| Percent Split: 66% | 48% |
| Percent Split: 75% | 50% |

Numerical prediction based on “msl\_trend\_ft\_century” class:

|  |  |
| --- | --- |
|  | Linear Regression |
| Cross-Validation: 5 | 58% |
| Cross-Validation: 10 | 59% |
| Cross-Validation: 20 | 60% |
| Percent Split: 50% | 51% |
| Percent Split: 66% | 50% |
| Percent Split: 75% | 50% |