Data mining with KNN and Random Forest

1. with data balance
   1. KNN – under sampling

|  |  |  |
| --- | --- | --- |
| Repeat 100 times | Max accuracy | K |
| Lab A | 80.65% | 17 |
| Lab A (Min-Max normalization) | 83.87% | 22 |
| Lab A (Z-score standard) | 88.71% | 10 |
| Lab B | 80.43% | 8 |
| Lab B (Min-Max normalization) | 78.26% | 1 |
| Lab B (Z-score standard) | 80.43% | 16 |

* 1. Random Forest – over sampling

|  |  |  |
| --- | --- | --- |
| Repeat 100 times | Average accuracy | Max accuracy |
| Lab A | 80.68% | 88.24% |
| Lab A (Min-Max normalization) | 75.18% | 80.15% |
| Lab A (Z-score standard) | 80.39% | 87.50% |
| Lab B | 74.33% | 82.76% |
| Lab B (Min-Max normalization) | 75.18% | 80.15% |
| Lab B (Z-score standard) | 80.39% | 87.50% |

1. without data balance
   1. KNN

|  |  |  |
| --- | --- | --- |
| Repeat 100 times | Max accuracy | K |
| Lab A | 81.82% | 10 |
| Lab A (Min-Max normalization) | 84.85% | 6 |
| Lab A (Z-score standard) | 84.85% | 14 |
| Lab B | 76.92% | 7 |
| Lab B (Min-Max normalization) | 78.85% | 12 |
| Lab B (Z-score standard) | 75.00% | 9 |

* 1. Random Forest

|  |  |  |
| --- | --- | --- |
| Repeat 100 times | Average accuracy | Max accuracy |
| Lab A | 82.85% | 86.87% |
| Lab A (Min-Max normalization) | 73.53% | 78.79% |
| Lab A (Z-score standard) | 83.39% | 85.86% |
| Lab B | 75.23% | 80.77% |
| Lab B (Min-Max normalization) | 76.87% | 86.54% |
| Lab B (Z-score standard) | 76.46% | 80.77% |

Note: All data have been cleaned.