1. **Clustering without PCA:**

a) k-Means

1. k-means with k = 2: SSE = 11823.4520395564

Confusion matrix (Predicted vs Actual)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  |  | Actual |  | |  |  | 0 | 1 | |  | 0 | 6 | 4 | | predict | 1 | 5 | 23 |   Accuracy=76.31% |  |  |  |
|  |  |  |  |

2. k-means with k = 3: SSE = 11113.6370513099

Confusion matrix (Predicted vs Actual)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Actual |  |  |
|  |  | 0 | 1 | 2 |
|  | 0 | 6 | 1 | 1 |
| predict | 1 | 4 | 10 | 3 |
|  | 2 | 1 | 8 | 4 |

Accuracy=52.63%

b) Hierarchical Clustering

1. hierarchical with k = 2:

Confusion matrix (Predicted vs Actual)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual |  |
|  |  | 0 | 1 |
|  | 0 | 26 | 11 |
| predict | 1 | 1 | 0 |

Accuracy=68.42%

2. hierarchical with k = 3:

Confusion matrix (Predicted vs Actual)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Actual |  |  |
|  |  | 0 | 1 | 2 |
|  | 0 | 17 | 11 | 8 |
| predict | 1 | 1 | 0 | 0 |
|  | 2 | 1 | 0 | 0 |

Accuracy=44.73%

1. **Clustering with PCA:**

a) k-Means

1. k-means with k = 2: SSE = 33.6727065522362

Confusion matrix (Predicted vs Actual)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual |  |
|  |  | 0 | 1 |
| predict | 0 | 7 | 7 |
|  | 1 | 4 | 20 |

Accuracy=71.05%

2. k-means with k = 3: SSE =30.5645148378693

Confusion matrix (Predicted vs Actual)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Actual |  |  |
|  |  | 0 | 1 | 2 |
| predict | 0 | 2 | 4 | 1 |
|  | 1 | 4 | 11 | 3 |
|  | 2 | 2 | 3 | 7 |

Accuracy=52.63%

b) Hierarchical Clustering

1. hierarchical with k = 2:

Confusion matrix (Predicted vs Actual)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual |  |
|  |  | 0 | 1 |
| predict | 0 | 26 | 11 |
|  | 1 | 1 | 0 |

Accuracy=68.42%

2. hierarchical with k = 3:

Confusion matrix (Predicted vs Actual)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Actual |  |  |
|  |  | 0 | 1 | 2 |
| predict | 0 | 18 | 11 | 7 |
|  | 1 | 1 | 0 | 0 |
|  | 2 | 0 | 0 | 1 |

Accuracy=47.36%

1. **Classification**

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Parameters** | **Training Accuracy** |
| Random Forest | numTrees=100, seed=1 | 100.00% |
| Boosting | weghtthreshold=100 | 5.26% |
| Bagging | numberIfIterations=100, seed=1 | 55.26% |
| J48 | confidenceFactor=0.25, minNumObj=2, numFolds=3, seed=1 | 44.74% |

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Parameters** | **Test Accuracy** |
| Random Forest | numTrees=100, seed=1 | 5.26% |
| Boosting | weghtthreshold=100 | 2.63% |
| Bagging | numberIfIterations=100, seed=1 | 2.63% |
| J48 | confidenceFactor=0.25, minNumObj=2, numFolds=3, seed=1 | 0.00% |