

CSE 508 Information Retrieval Assignment 3

I. 50:50 Split

Accuracy: 95.04 %

Confusion Matrix:

| | | Predicted | | | | |
|--------|---|-----------|------|------|------|--------|
| | | 0 | 1 | 2 | 3 | 4 |
| Actual | 0 | [[450. | 1. | 36. | 13. | 0.] |
| | 1 | [6. | 490. | 4. | 0. | 0.] |
| | 2 | [6. | 1. | 478. | 9. | 6.] |
| | 3 | [10. | 0. | 15. | 473. | 2.] |
| | 4 | [1. | 0. | 5. | 9. | 485.]] |

II. 70:30 Split

Accuracy: 96.33 %

Confusion Matrix:

| | | Predicted | | | | |
|--------|---|-----------|------|------|------|--------|
| | | 0 | 1 | 2 | 3 | 4 |
| Actual | 0 | [[284. | 0. | 9. | 6. | 1.] |
| | 1 | [1. | 295. | 3. | 0. | 1.] |
| | 2 | [5. | 0. | 284. | 7. | 4.] |
| | 3 | [2. | 0. | 4. | 294. | 0.] |
| | 4 | [4. | 0. | 5. | 3. | 288.]] |

III. 80:20 Split

Accuracy: 96.39 %

Confusion Matrix:

| | | Predicted | | | | |
|--------|---|-----------|------|------|------|--------|
| | | 0 | 1 | 2 | 3 | 4 |
| Actual | 0 | [[192. | 0. | 5. | 2. | 1.] |
| | 1 | [1. | 196. | 0. | 2. | 1.] |
| | 2 | [2. | 0. | 191. | 5. | 2.] |
| | 3 | [5. | 0. | 2. | 190. | 3.] |
| | 4 | [0. | 0. | 3. | 2. | 195.]] |

IV. 90:10 Split

Accuracy: 98.0 %

Confusion Matrix:

| | | Predicted | | | | |
|--------|---|-----------|------|-----|-----|--------|
| | | 0 | 1 | 2 | 3 | 4 |
| Actual | 0 | [[95. | 0. | 1. | 3. | 1.] |
| | 1 | [0. | 100. | 0. | 0. | 0.] |
| | 2 | [0. | 0. | 99. | 0. | 1.] |
| | 3 | [1. | 0. | 2. | 96. | 1.] |
| | 4 | [0. | 0. | 0. | 0. | 100.]] |

We can clearly see that with the increase in training data, accuracy increases.