**CSE3506 – ESSENTIALS OF DATA ANALYTICS**

**EXPT NO: R. Harini**

**DATE:** 17/3/2021

**NAME:** R. Harini

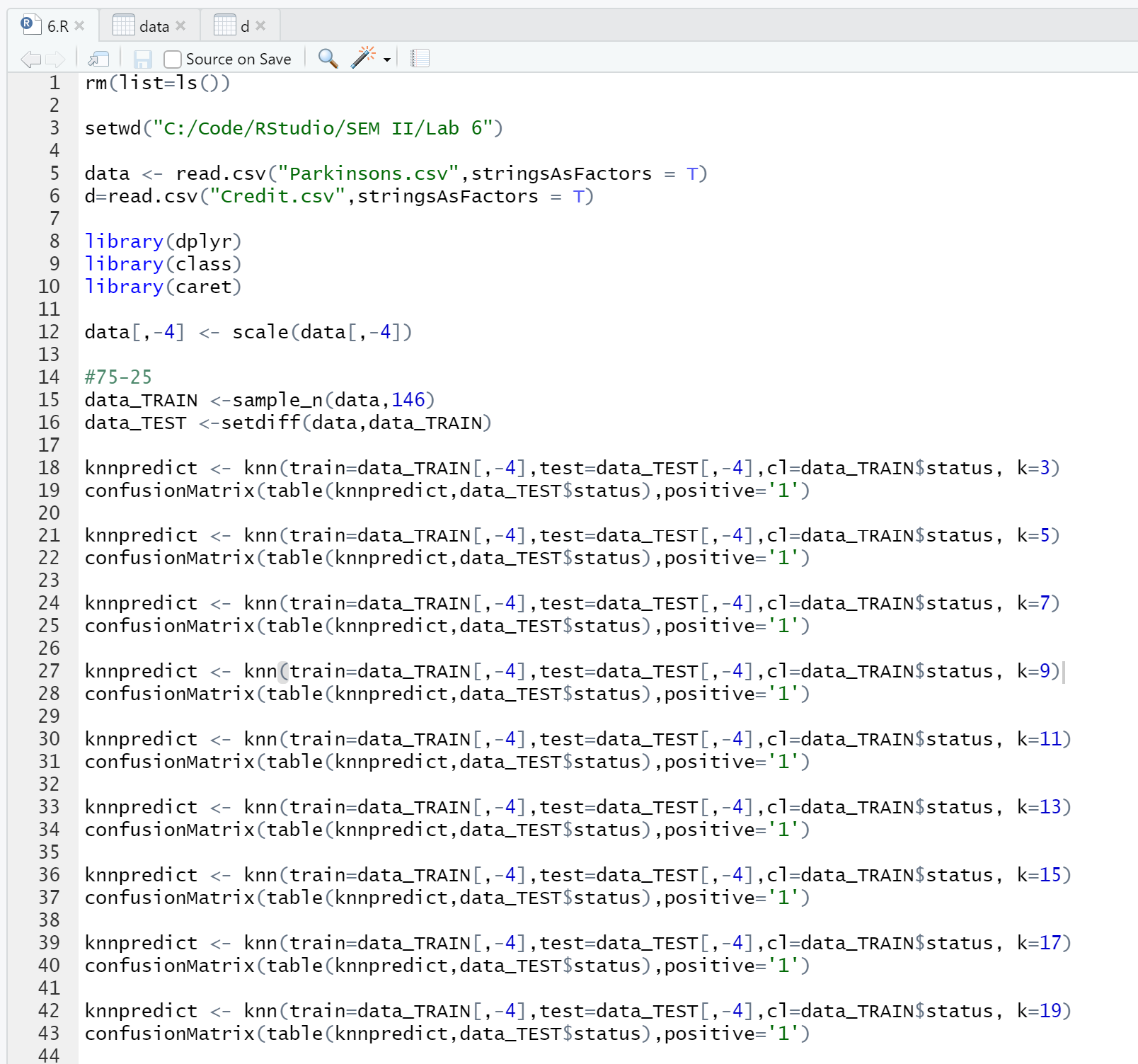
**REG. NO.:** 18BCE1010

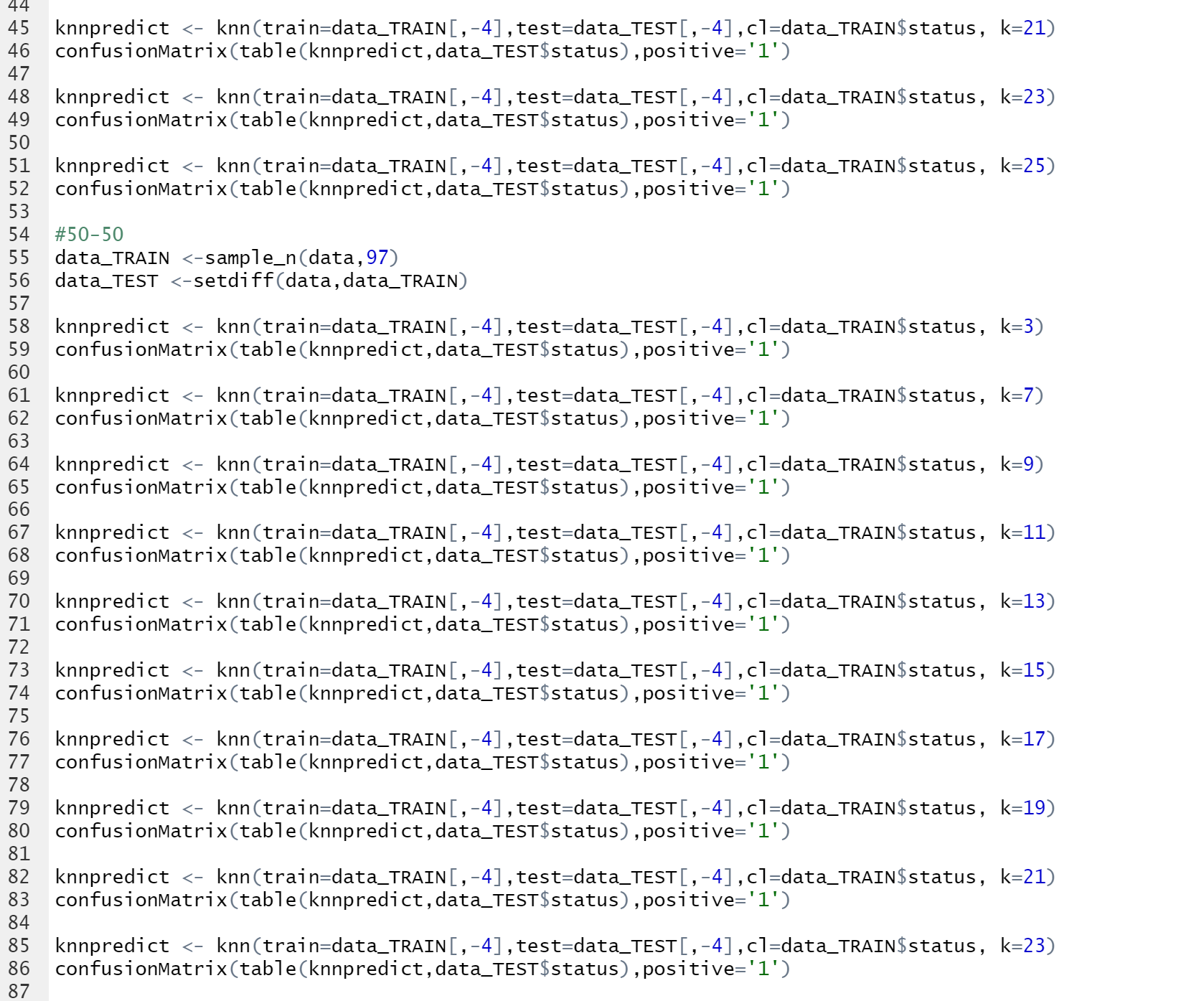
**K- NEAREST NEIGHBOR**

**AIM**

Understanding the K-Nearest Neighbor algorithm on ‘Credit.csv’ file and then performing the same algorithm on Parkinsons.csv for different values of ‘k’

**MATLAB PROGRAM**

****

****

**OUTPUT**

Total number of observations = 195

Number of training observations = 146 (75%)

Number of testing observations = 49 (25%)

|  |  |  |  |
| --- | --- | --- | --- |
| K Value | Accuracy | Sensitivity | Specificity |
| 3 | 0.9388 | 0.9459 | 0.9167 |
| 5 | 0.898 | 0.9189 | 0.8333 |
| 7 | 0.9184 | 0.9459 | 0.833 |
| 9 | 0.9388 | 0.9730 | 0.833 |
| 11 | 0.9184 | 1.000 | 0.6667 |
| 13 | 0.8367 | 0.9459 | 0.5000 |
| 15 | 0.8571 | 0.9459 | 0.5833 |
| 17 | 0.8571 | 0.9730 | 0.5000 |
| 19 | 0.8776 | 1.000 | 0.5000 |
| 21 | 0.8776 | 1.000 | 0.5000 |
| 23 | 0.8776 | 1.000 | 0.5000 |
| 25 | 0.8776 | 1.000 | 0.5000 |

The highest accuracy is obtained by **k=3** and **k=9** with an accuracy of **0.9388**

Number of training observations = 97 (50%)

Number of testing observations = 98 (50%)

|  |  |  |  |
| --- | --- | --- | --- |
| K-Value | Accuracy | Sensitivity | Specificity |
| 3 | 0.8469 | 0.8904 | 0.7200 |
| 5 | 0.8367 | 0.8767 | 0.7200 |
| 7 | 0.8776 | 0.9178 | 0.7600 |
| 9 | 0.8878 | 0.9315 | 0.7600 |
| 11 | 0.8776 | 0.9178 | 0.7600 |
| 13 | 0.8878 | 0.9589 | 0.6800 |
| 17 | 0.7857 | 0.9315 | 0.3600 |
| 19 | 0.7857 | 0.9452 | 0.3200 |
| 21 | 0.8061 | 0.9726 | 0.3200 |
| 23 | 0.7959 | 0.9726 | 0.2800 |
| 25 | 0.8061 | 0.9863 | 0.2800 |

The highest accuracy is obtained by **k=9** and **k=13** with an accuracy of **0.8878**