**Quiz 4 - K Nearest Neighbours**

Consider using the data set “Smarket.csv” to predict direction using Lag1 through Lag5. The data was split into pre-2005 and post-2005 subsets to build a KNN model. Refer to the photo provided for all subsequent questions.

A white screen with blue text

Description automatically generated

|  |  |
| --- | --- |
| **Question 1**  Which dataframe contains the testing data? | |
| A | market |
| B | market1 |
| C | market2 |
| D | None of the above |

|  |  |
| --- | --- |
| **Question 2**  What does the object “pred” contain? | |
| A | The predicted outcomes for data points before 2005 |
| B | The predicted outcomes for data points in 2005 |
| C | The predicted outcomes for all data points |
| D | The real outcomes for all data points |

|  |  |
| --- | --- |
| **Question 3**  Consider the first data point where the predicted response for it in pred is Down. What is the maximum number of neighbours for that point whose true outcome is "Up"? | |
| A | 1 |
| B | 2 |
| C | 3 |
| D | 4 |

|  |  |
| --- | --- |
| **Question 4**  Suppose you wanted to evaluate the performance of your KNN model with a confusion table. What inputs are required? | |
| A | Predictions on the training data & actual outcomes of the testing data |
| B | Predictions on the testing data & actual outcomes of the testing data |
| C | Predictions on the testing data & actual outcomes of the training data |
| D | None of the above |

|  |  |
| --- | --- |
| **Question 5**  Select all the best practices when building a KNN model. | |
| A | Scale quantitative data |
| B | Use the first 80% of data points as your training data |
| C | Use odd values of K |
| D | Include all available quantitative & categorical features |