

CISC 251 - Sheet 3

AMY BRONS - 20252295
SEPT 29 2021

QUESTION 1:

When looking at how to experiment with the k-Nearest predictor, we are attempting to classify the election outcome by varying various parameters within the predictor.

When experimenting I noticed a couple interesting points:

1. Accuracy was higher when the neighbours considered was higher.
2. Weight of neighbours considered has very little improvement, but is slightly more accurate.
3. If neighbours distance was not considered, the number of considered neighbours would have to be 1 to get 100% accurate prediction.

To demonstrate, here is a view when there is only one neighbour is considered but the weight of neighbours is considered:

Confusion Matrix - 3:12 - Scorer

File

Hilite

Icdiscrete ...	H	M	L	
H	7	10	1	
M	11	15	4	
L	1	6	2	

Correct classified: 24

Wrong classified: 33

Accuracy: 42.105%

Error: 57.895%

Cohen's kappa (κ): 0.017%

Here is where distance is considered, and neighbours is 29:

Confusion Matrix - 3:12 - Scorer

File	Hilite			
Icdiscrete ...	H	M	L	
H	9	8	1	
M	5	24	1	
L	2	7	0	

Correct classified: 33


Wrong classified: 24

Accuracy: 57.895%

Error: 42.105%

Cohen's kappa (κ): 0.228%

Here is where distance is not considered, and neighbours is 32:



Confusion Matrix - 3:12 - Scorer

File	Hilite			
Icdiscrete ...	H	M	L	
H	9	9	0	
M	5	24	1	
L	2	7	0	

Correct classified: 33

Accuracy: 57.895%

Cohen's kappa (κ): 0.219%

Wrong classified: 24

Error: 42.105%

QUESTION 2:

Most predictions are quite difficult, as the K Neighbour needs to be adjusted and tuned for each output. It is quite difficult to score an accuracy over 50%, and therefore it is hard so call this method 'clear-cut'. In the csv this can be seen if we scroll and find the exact data that would equal this closeness, but unfortunately this would not be able to be seen in KNIME without additional processing, particularly partitioning, which would require dates included in the uselection csv file.