

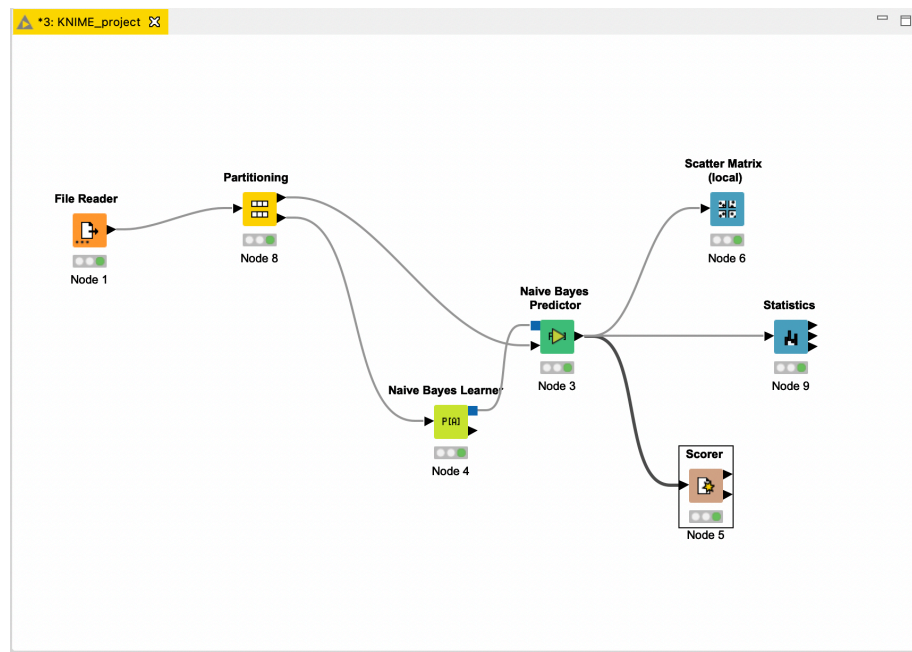
# CISC 251 - Exercise Sheet 2

AMY BRONS - 20252295

SEPT 23 2022

## Question 1:

Using the uselection.csv file in KNIME, I was able to set up a workflow that utilizes the Naive Bayes classifier. Here is a breakdown of the starting workspace:



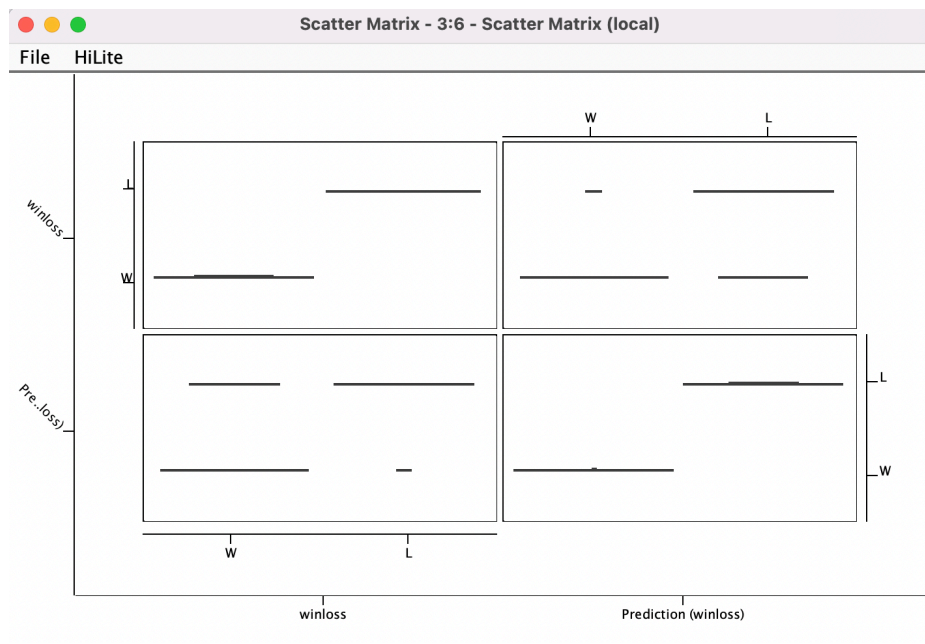
After that, conclusions can be drawn out of this model. For example I was able to extract a confusion matrix that analyzes the win/loss prediction per speech, and classification accuracy:

Confusion Matrix - 3:5 - Scorer			
File	Hilite		
Prediction ...	W	L	
W	113	12	
L	69	107	

Correct classified: 220	Wrong classified: 81
Accuracy: 73.09%	Error: 26.91%
Cohen's kappa (κ): 0.48%	

Also I was able to extract a scatter matrix of how the predicted won/loss compared to the actual win/loss:



Finally, I experimented with binning in KNIME. I made an interactive histogram with 'the' as the binning element. I compared this to the words 'america' and 'president'. Additionally, to make the data easier to read on our scale, I reduced the bin number to 5. Here is the result of that data

