

## Homework 2 – Performance Assessment (updated 1/16/18)

### Part I

You have just hired 10 undergraduates to label a large number of Facebook profile pictures for gender and ethnicity. You provide a written codebook and send each of them off with the same 200 images to label. A week later, you find that average agreement is above 80%.

1. Does this level of agreement indicate high reliability? Explain.
2. What does reliability tell you about your project?
3. What does Krippendorff suggest you should be asking in evaluating agreement?
4. Provide three specific examples of how you might assess/demonstrate the validity of your facebook coding effort?

### Part II

In supervised learning, the standard practice is to report agreement between the true or gold standard labels and the algorithm's predictions. Three scores are typically reported. Accuracy is overall correctly predicted percentage. Precision is the percentage of correct predictions for a given category. Recall is the percentage of true cases in a category that are correctly predicted.

In the following table, the rows indicate the true cases while the columns represent the predictions

5. What is the accuracy?
6. What is the precision for defense
7. What is the recall for defense?
8. What is the false positive (type I error) rate
9. What is the false negative (type II error) rate?
10. What is the expected accuracy in this example?

	defense	education	health
defense	30	3	5
education	4	19	2
health	7	6	27

*#Here's one Confusion matrix example in R (better example? Let me know!*

*#<https://www.rdocumentation.org/packages/caret/versions/3.45/topics/confusionMatrix>*

```
install.packages('caret')
library(caret)
install.packages('e1071', dependencies=TRUE)
library(e1071)
```

```
mypred = c('a','a','b','b','c','c','b','c') #could also be integers, words, whatever
```

```
mytrue= c('b','a','b','b','a','a','b','c')  
confusionMatrix(factor(mypred), factor(mytrue))
```

*#factor means that the values in the vector are converted to characters (predictions don't have to be numeric)*

*#In the output, accuracy is the overall percent correctly predicted;  
#sensitivity is precision; pos pred value is recall  
#a positive Kappa value indicates above expected accuracy*

### Part III

Keep testing your knowledge of python using `alottapythonexamples.py` and options in `quanteda`

Also, if you are using `spyder` and want to open plots etc in their own window, google it. After you make the changes you have to restart `spyder` for it to work

Start thinking about your project proposal – come see me to discuss!