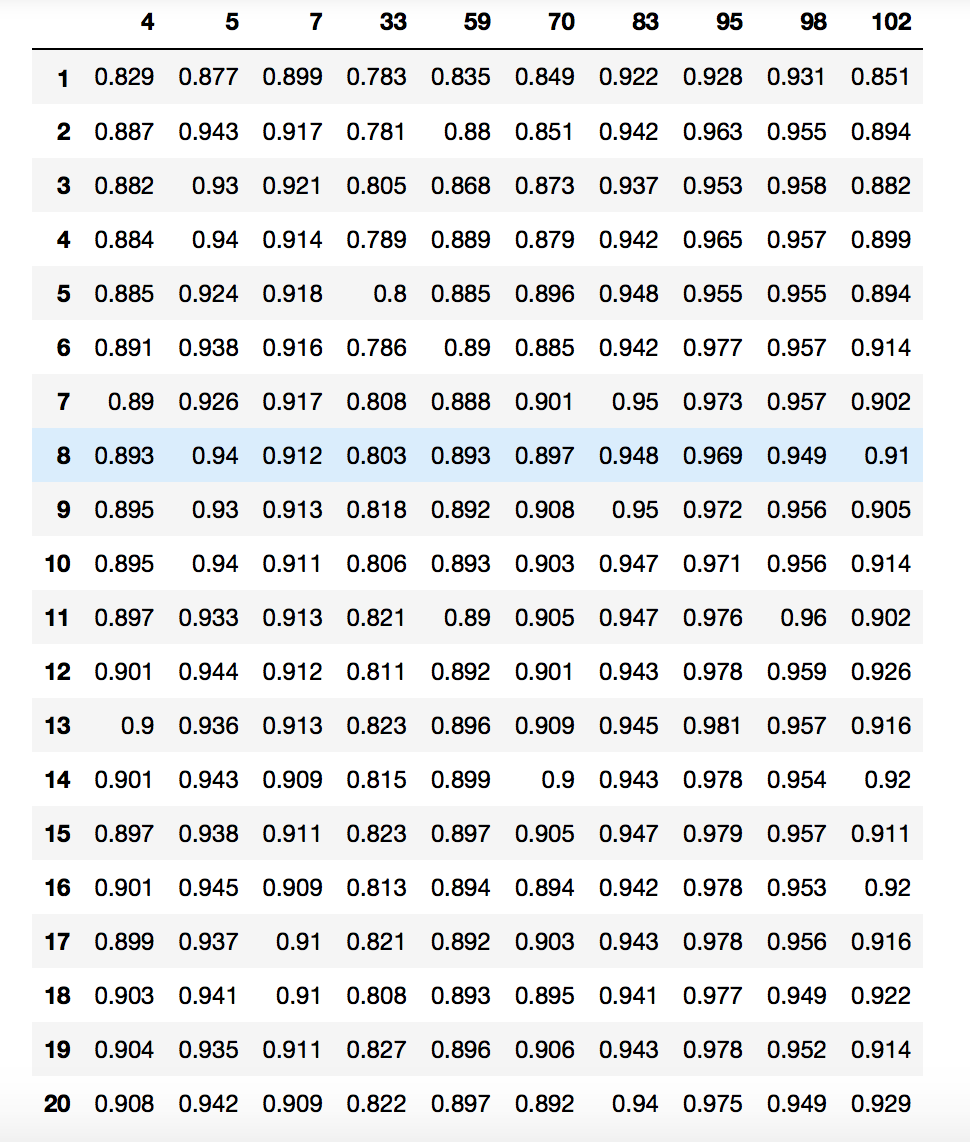
# K-Nearest-Neighbors and Naive Bayes

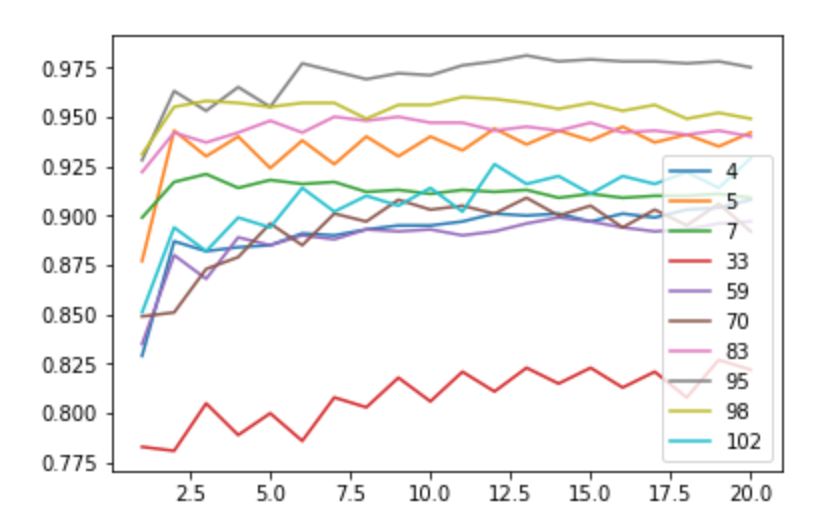
## Classification with kNN

Finding the value of k for each label was done by sweeping across all labels we were concerned with and all values of k from 1 to 20. Training (on the train set) and predicted (on the validation set) based on each label and then scoring the predictions. This yielded the matrix shown here:



This matrix shows the score (0 – 1 with 0 being all predictions are wrong and 1 being all predictions are correct) of each label with each value of k. (the rows represent values of k and the columns represent the id of the labels).

For each label, a graph of the score vs the value of k was then produced using df.plot().



From this graph and table, an appropriate value of k was found for each class. These were then recorded in the ‘ks.txt’ file.

## Classification with NB