1. For all questions: Screenshot of each methods/classes (code) you wrote (for all question), simple explanation of each method (if you include comments in the code, that would be sufficient as a snapshot)

Please see the HTML or jupiter notebook file I submitted

2. question 3 and 4: The plot of your metric with different K (let's say five different K are tried) (either 5 curves with x-axis as ten folders, or one curve with x-axis with different K)

Please see the plot I incorporated in my submission

3. question 5: K chosen, your confusion matrix and overall accuracy.

K = 3, best accuracy overall is 81% with SVM on 10-fold CV

4. For the report on question 7 you will need to include all the snapshots of the processes of rapid
miner and add short comments on your learning outcomes from this exercise.

I didn't use rapid miner for ensemble learning, instead I just implemented it in python, because

(1) The time limit, I'm more fluent to do ensemble learning in python. Yet for rapid miner i need to first save my vectors for texts, then do all those works which requires more time to put my hands on.

(2) People today in industry are using python and Scala for offline training, and GO for online deployment.