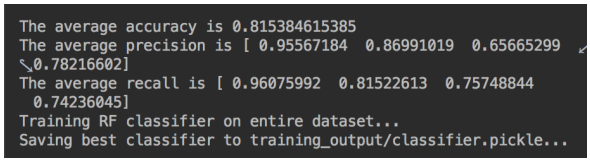
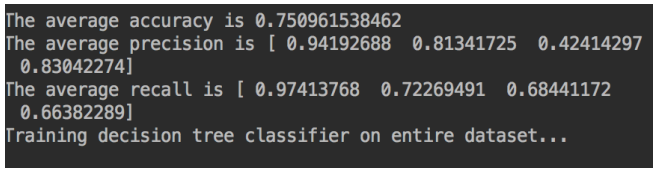
Write-up

Answer the following questions.

1. Please submit your CSV files with audio data to the repo. You may have to use git add --force in order to add them as the repo is configured to ignore .csv extensions.
2. Describe the best classifier model you used and report its average precision, recall, and accuracy over the 10 folds. Also describe which features you implemented on top of the formants and pitch data (mean, variance, etc.).

   We chose Random forest tree because it performed the best





(RF) (DT)

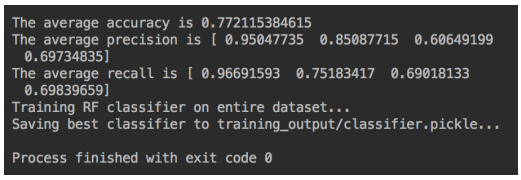
We also implemented mean, variance and energy over the window as well as Delta and Delta-Delta. The final classifier didn't use Delta-Delta, since it negatively affected our results.

3. Discuss the changes you made to improve accuracy. How well did things work? What worked best? Give concrete accuracy / precision / recall numbers so we have a sense of how much your classification improved from the baseline.

1.Speaking for a longer time gave the classifier more data to train with

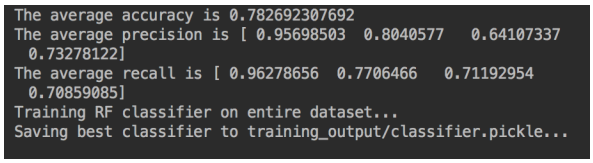
2. Testing in an area without significant noise or interfering sounds

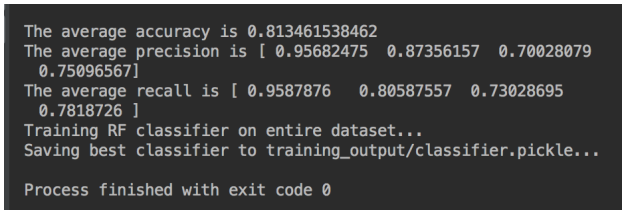
3. Increase value of N when calculating delta and delta-delta



(RF N=2) (RF N=10)

4. Add more features (this would help DT more than RF)



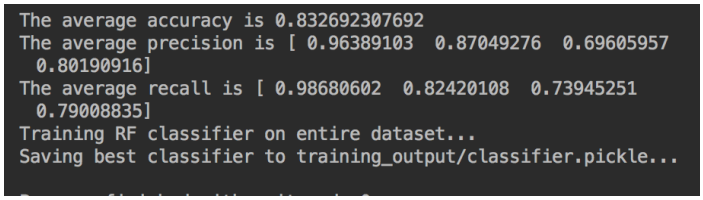


(RF N =10, added mean,variance, energy,delta-delta) (same but without delta-delta)

4. Increase # of bins, increase DT depth(but we didn't use DT in the end), increase n_estimator number(RF) would help a little bit.

5. In the beginning, we only had a 70% accuracy, what helped the most was changing the # of bins and value of N. Also, we recollected one group member's data when we found his precision and recall was significantly lower than the other members'. (we don't have those early screenshots but after changing all the values mentioned above, our accuracy reached 80%)

6.An interesting thing that happened was when we removed Delta-Delta, accuracy increase. When we removed Delta and Delta-Delta, it even increased to 84%!! However, the real-time performance was significantly worse.



(RF without delta and delta-delta)

4. How did you collect your data? Describe the acoustic environments and any other factors that may have an effect on your classifier. How do you account for variation?

The no speaker data was collected in W.E.B Du Bois Library. The phone remained in the pocket as one person walked around at the Lower Level for 3 minutes. At first, the other 3 speakers' data are from speeches by Barack Obama, Donald Trump and Hillary Clinton. This data was very high-quality, and made the precision of the classifier very high. However, since this was not allowed, we instead collected data based on our own voices for about 4 minutes each. One piece of data was collected at a house, while the great majority of the data was collected in Marcus hall at two places: At the entrance and in the cafeteria. The entrance had some people making footsteps as well as opening/closing doors, while the cafeteria had a refrigerator that made a constant drone. These ambient sounds likely generated some noise in our data, but the data came out rather smoothly regardless.To account for variation, in the end, we recollect data at marcus hall using the same phone while reading similar article.(surrounding is pretty quiet) When collecting data, we each maintained the same distance from the microphone.

5. Describe another way (besides the one mentioned in the Background section) that speaker identification can be used for health analytics.

1.Test a person's stress level, whether they're having an anxiety attack

2. Test whether they are depressed or not. Would need a lot more data and wouldn't be the primary method of determination, but could be a hint.

3. Respiratory disease/conditions

6.Josh claims that one's vocabulary might be informative enough to distinguish speakers; for instance, he might say "wicked" once in awhile, but his mom still thinks of witches when she hears the word. Assuming that speech recognition tools are as good as we are at hearing, what problem might you still face when trying to implement this feature for speaker recognition? (There is an extra credit part where you can try it!).

Using speaker recognition based on vocabulary has the problem of people saying different things when they talk to different people. The same person at work versus at home versus on vacation could all have very different speech patterns, and could make the classifier think they are completely different people. Besides speaking patterns, we also need to analyse a person's voice features in order to recognize a speaker. Vocabulary could be an addition to identifying a person, but doesn't seem distinctive enough to warrant identifying a person on its own.