There are two general areas that lend themselves two future work: 1) classifier generalization improvement and 2) application implementation.

For classifier generalization, the next logical step in improvement would be to add a layer of classification that attempts to classify the training dataset that the input is most similar to. The reason for this is that in tuning the classifier, we noticed a ~5% increase in both accuracy and F-score of the ensemble classifier when the classifier was given an additional feature of the type of data on which it was classifying (e.g. ‘Twitter Data’, ‘SMS Data’ etc.). While a potential user is fully capable of inputting the context in which the classifier is currently classifying, having another layer in which the ensemble classifier uses machine learning to detect the learned dataset to which the input is most similar would make for a more efficient and seamless classifier.

For application implementation, the authors had the original intention of plugging directly into the Facebook messenger API to help give a gauge of peer’s mental health. This would require a few pieces of extension that were discussed, but not implemented. While we did create code to read the inputs of the user in real time and classify them, all of the inputs are classified as unique data points and thus a graph of a user’s sentiment over time would seem stochastic even though it is likely that a normal individual’s emotional state would not change randomly from second to second. Thus a low pass filter on the emotional state of an individual – as determined by their sentiment – over time would help improve the accuracy of the classification in a real-time context. Also, because we are assuming an individual’s emotional state is steadier for smaller time steps between inputs, the time step in-between inputs should be considered. One plausible candidate for doing such a calculation would be to classify an individual’s emotional state based on the sentiment of their current input plus the change in their sentiment over time multiplied by some constant.