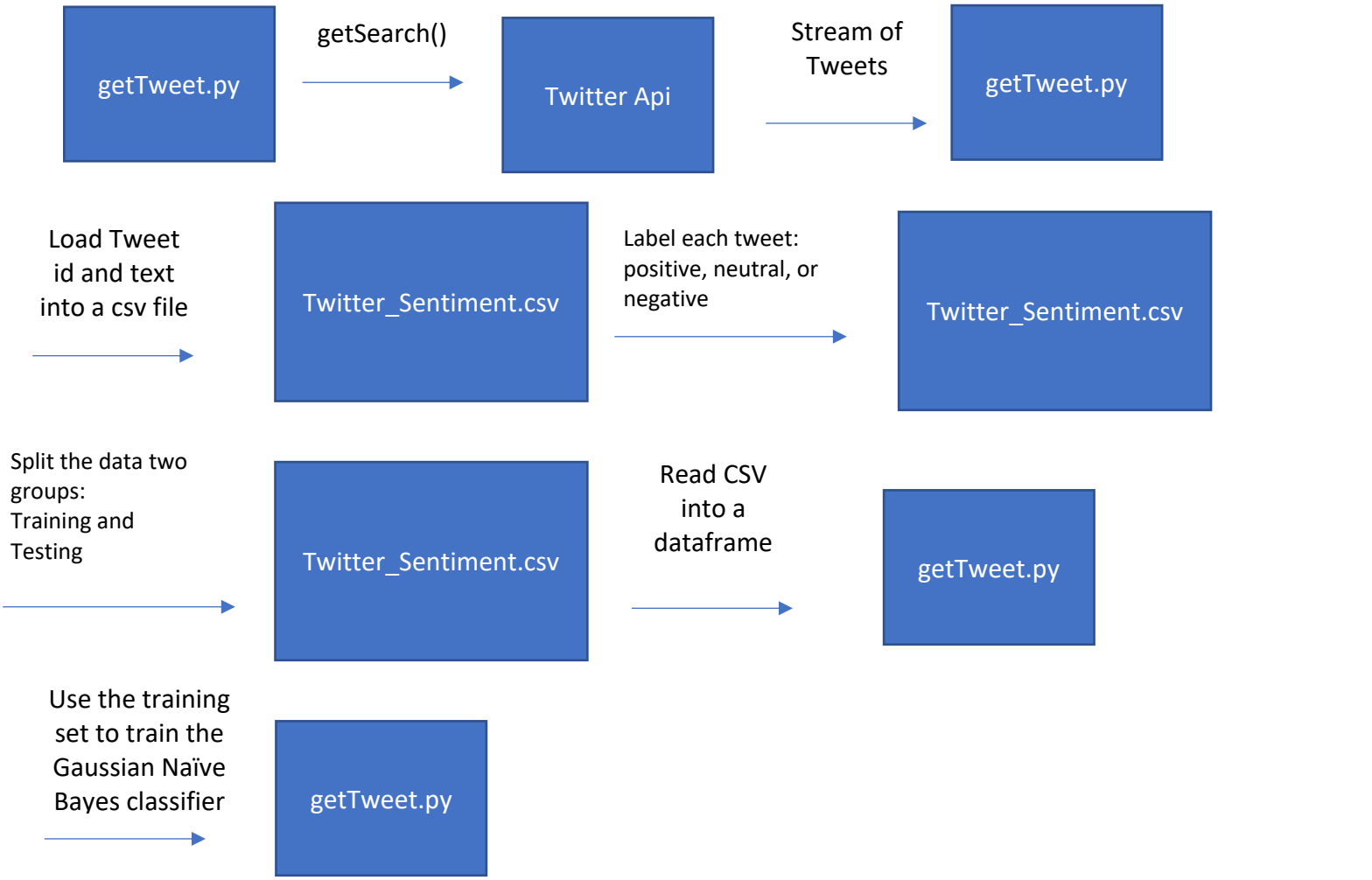
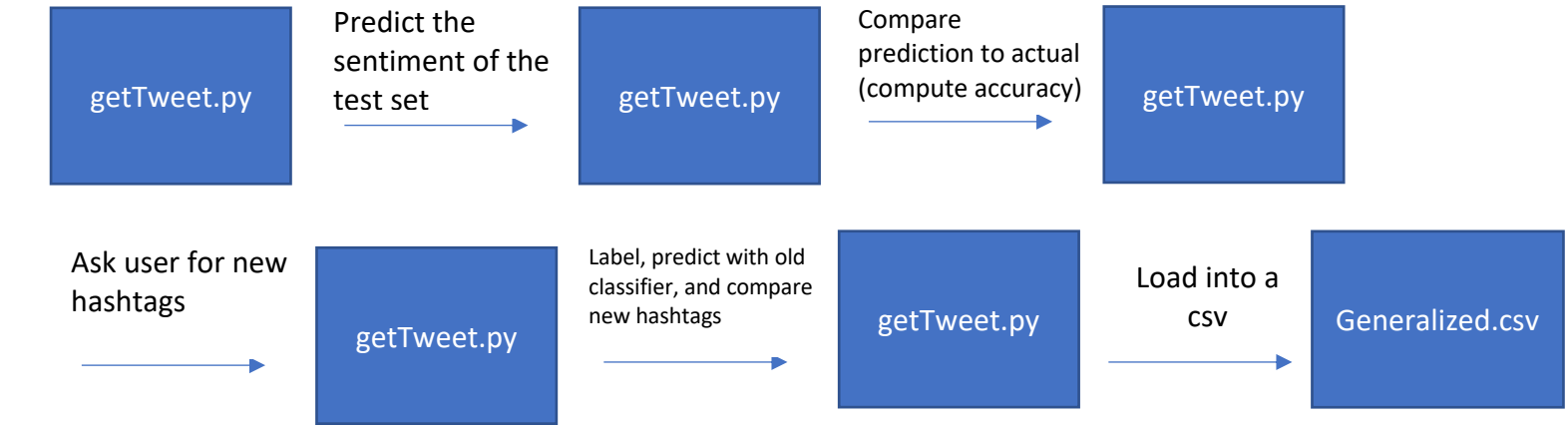
Stage 1: Collecting Training Tweets



Stage 2: Predictions and Accuracy



2. Explain how you tuned your model to achieve over 80% accuracy

For me to achieve over 80% accuracy I had to play with what tweets I used a lot. I found using hashtags that were too monotone, ie not having enough opinionated words, would present mostly neutral tweets, which made the classifier think almost everything was neutral. Once I found three hashtags that got me around 70%, I went back through the tweets and relabeled them, trying to make sure they were as correctly labeled as I could. From there on it was about selecting the right tweets to train, so that the classifier had a wide range of positive, neutral, and negative labeled tweets for training.

3. My classifier worked fairly well on the initial hashtags, with an accuracy of 84.84%, but averaged around 50% on any other group of three tweets. This tells me that the classifier can predict the sentiment of things like the training set, but cannot accurately predict what the sentiment classification of unrelated tweets is likely due to the intricate duality of the English language.