Weather & Twitter Project Documentation

Originally we started to take forecasts for 3 locations: London, Manchester and Glasgow.

Had to settle for just London and Glasgow because there were no good observations the area near Manchester surprisingly. Only Liverpool and by this point we had already been taking the forecasts from Manchester.

We took the observations with the aim of comparing forecasts from 5days, 3days, 1day, xHrs ago to the actual forecast.

We started downloading tweets from London and Glasgow on the 8th of December and putting them in MongoDB. Chay had to use iJson because the JSON files couldn’t be loaded into Python because they were too large. The aim was to filter all the tweets by weather words and then run those tweets through the algorithm. Those that were left could then be analysed through Python and put into a pandas dataframe.

We put all the tweets into MongoDB, from there we ran a for loop to add a key called ‘W1’ which returned ‘Y’ when weather related and ‘N’ when not. We then ran the algorithm over this list which left us with tweets which should in theory be predominantly weather related.