***17/02/2020 - MONDAY***

* Members and Roles assigned in mutual agreement:
  + Chair – Cian
  + Recorder – Aishwarya
  + Accountant – George
  + Verifier – Sameer
  + Ambassador – Shravani
* Topics considered by the group – “Does sentiment of public transport users in Ireland correlate with categories implicit within Met Eireann textual forecast data?” and “Within one country, and during fixed temporal periods, to what extent do pairwise twitter posting similarity measures correspond to pairwise post geo-location distances”?
* Topic finalised by the group – Sentiment analysis on weather data from Twitter by public transport users in Ireland
* Current step
  + Focus on finding papers based on sentiment analysis based on weather and twitter.
* Still in discussion
  + Geo-location / locations considered in Ireland
  + Time period of tweets

***20/02/2020 - THURSDAY***

* Discussed what needs to be done
  + Sentiment analysis
  + Public transport users
  + Correlation with weather data
* Idea is start of simple by just trying to perform sentiment analysis of only tweets regarding weather and incorporate with public transport users.
* To-do:
  + Aishwarya and George – Grab tweets
  + Cian – Get weather data and check the LateX template
  + Shravani and Sameer – Research on sentiment analysis algorithms

***25/02/2020 - TUESDAY***

1. <https://www.nltk.org/api/nltk.sentiment.html>

NLTK.sentiment package – A SentimentAnalyzer is a tool to implement and facilitate Sentiment Analysis tasks using NLTK features and classifiers

1. <https://twitter.com/meteireann>

MetEireann’s twitter data

1. <https://towardsdatascience.com/creating-the-twitter-sentiment-analysis-program-in-python-with-naive-bayes-classification-672e5589a7ed>

Creating the Twitter Sentiment Analysis Program in Python with Naive Bayes Classification

1. <https://archive.org/details/twitterstream>

A simple collection of JSON grabbed from the general twitter stream, for the purposes of research, history, testing and memory. This is the "Spritzer" version, the most light and shallow of Twitter grabs.

1. <https://archive.org/details/archiveteam-twitter-stream-2018-10>
2. <https://github.com/sangaline/wayback-machine-scraper>

repository consists of a command line utility that can be used to scrape data as it appears in archive.org’s Wayback Machine.

***03/03/2020 – TUESDAY***

* Discussed regarding the draft for the report
  + Introduction
  + Related Work/Background/Literature Reviews
  + Methods
  + Results
  + Conclusions
* Tasks –
  + Aishwarya and George – To write the intro and Related work
  + Shravani and Sameer – To try sentiment analysis on the data
  + Cian – Data exporting + weather categories

***08/03/2020 – TUESDAY***

* RESEARCH QUESTION -Can we identify a correlation between Implicit categories of Textual Weather data and Sentiment Analysis of Twitter data for the inhabitants of Ireland?
* Rename related work to literature review
* Tasks –
  + Aishwarya – Research question
  + Shravani – Update box plots and information regarding it into the essay
  + George – Conclusions and Results
  + Cian – coagulate everything

***30/03/2020 – MONDAY***

* Cian
  + weather categories
  + counting the number of times the category is repeated in the tweet
  + how to weight them?
  + subjectivity and sentiment of the weather
* shravani
  + behaviour on weekdays and weekend
* data we have is sparsed?
* met data is sparsed
* other data we have for everyday
* but keep in account that there might be days with no data.
* forcast data is predictions
* particular day and not time at which

***10/04/2020 – FRIDAY***

* Expanding the dictionary, Replacing smileys
* Sentiment scores will be affected based on the mapping
* Several functions for pre-processing
* Loading all month’s data for the year 2018 – this is all weather forecast data.
* Combined all of it into a df
* 0 – Sunday and so on for the weeks of the days.
* Depends on total number of tweets found in a daye
* Cian – get the total observed number of tweets under each category and see if there is a comparison between expected counts of number of tweets under each category.

We have observed certain number of pos, neg and neutral tweets. If there is no correlation, then the null hypothesis becomes true that there is no correlation between the sentiment of the tweets with the weather.

* Pie charts – wet and cold the negative tweets are maximum
* **Overall pos and neg sentiments need to be related to total number and percentages of tweets, and compare observed and expected values based on some threshold value**
* Word cloud for pos, neg and neutral words also plotted.
* Is there a correlation between weather forecast and sentiment on twitter – you look at the app and see the weather and tweet something based on the mood. We are not looking at the actually weather data – POSSIBLE FUTURE WORK
* Referencing needs to be changed into a separate section. Follow the style that he uses in the handout – Cian will do the bibliography and significance measures.

***12/04/2020 – SUNDAY***

* **While other** studies have shown we can correlate.
* **Forecast** doesn’t necessarily because of multiple reasons - forecast is not always correct