

2. Project proposal (1 - 2 pages):

In 1 week, please have a chat with your team to choose the machine learning project ideas of interest. Please prepare 1 to 2 page project proposal presenting project title, datasets, timeline and the team members' contribution in percentage.

Please have a chat with the teaching team (me or our tutors) about the project idea to see whether it is a good fit as a final project for CAB420. You can submit the project proposal to the final project submission link. The final project submission link will appear after you sign up your group. There will be no mark for the project proposal.

<http://help.sentiment140.com/for-students/>

<https://github.com/fnielsen/afinn/blob/master/afinn/data/AFINN-en-165.txt>

Project Title

Twitter sentiment analysis

Team Members

Matthew Lord - n9743031

Angus Nobbs -n9736077

Sebastian Muir-Smith -n10199721

Joseph Grech -n9734902

Project Objective

The purpose of the project is to read a twitter tweet and determine whether it is 'positive' or 'negative'. This will be accomplished through using training data in the dataset to train a learner to read this text and classify it using a Naive Bayes classifier. Our project will use the AFINN word list that provides polarity of words as either positive or negative for a base line.

Datasets

The twitter datasets have been acquired from a Stanford university class project, in the project the class collected data and labelled it based on a scale of 0-4, with 0 being negative, 2 being neutral and 4 being positive. The twitter Search API was used and the data was collected through the use of keywords. Additionally, some further information such as the date the tweet

was posted, the user who tweeted and the text of the tweet saved in a csv file. This provides useful training data for the project.

References