CM4105 Honours Project – Literature Review

Jess McGowan, 1703725

## 1.0 Introduction

General description of topic

Machine learning is the practice of training a computer to make predictions, based on large-scale analysis of the past. It has a number of applications, especially regarding text-based analysis.

Highlight its importance

In particular, sentiment analysis of text is an area with a wide range of uses, from marketing analysis to gauging the overall mood of the public through analysing social media posts. Due to its public nature, Twitter is an effective website to use in order to get large amounts of data, with each post being limited in size to 280 characters (previously 140 characters).

A review of the literature is valuable in understanding important aspects of your research area

The literature covers some different approaches in how to carry out this analysis, and how it can be applied to different areas such as medical research, customer satisfaction, and politics. There are also a number of ways to visualise the findings.

Clarify the scope of the overall organisation of the idea

This review will focus on Twitter based sentiment analysis, and how it can gauge public sentiment effectively.

## 2.0 Machine Learning

Machine learning is the study of how computers can improve through experience without requiring human input (i.e. learning automatically) through the use of statistical analysis (Jordan & Mitchell, 2015). As such, complex analyses of big data can be processed in a much shorter period of time, compared to how long the process would take to complete manually.

### 2.1 Natural Language Processing

Natural Language Processing (NLP) is the area of research aimed at investigating how computers can learn how to understand natural language in text or speech (Chowdhury, 2005), and as a result, can build computer systems that can use NLP to complete tasks, as opposed to requiring input in a specific programming language.

### 2.1.1 NLTK

### 2.1.2 spaCy

### 2.2 Sentiment Analysis

As part of natural language processing, sentiment analysis is used to detect whether the sample text is positive, negative or neutral. This analysis is based on the polarity of the words used, which comes from a lexicon. The lexicon can either be a general one, or it can be specifically designed for an analysis where common terms can have a different sentiment polarity compared to general usage.

## 3.0 Twitter

Twitter is a popular social media network, with a focus on microblogging. It launched in 2006.

### 3.1 Twitter Sentiment Analysis

There have been a number of previous studies regarding sentiment analysis of Tweets (Agarwal, et al., 2011), (Go, et al., n.d.), (Hao, et al., 2011), (Kouloumpis, et al., n.d.) amongst others.

## 4.0 Web Data Visualisation Techniques

(methodological – different libraries/toolkits)

### 4.1 D3.js

## 5.0 Mainstream Media Impact on Social Media

(theoretical)

## 6.0 Summary and Project Objectives

"Can machine learning algorithms measure the impact of mainstream media effects on social media sentiment?"

# References

Agarwal, A. et al., 2011. *Sentiment Analysis of Twitter Data,* s.l.: Association for Computational Linguistics.

Agrawal, R., Anirudh, K., Dai, X. & Andres, F., n.d. Challenges and Opportunities with Big Data Visualization.

Allcott, H. & Gentzkow, M., 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives.*

Anon., n.d. Twitter\_and\_Society\_%282014%29.

Atefeh, F. & Khreich, W., 2015. *A SURVEY OF TECHNIQUES FOR EVENT DETECTION IN TWITTER,* s.l.: s.n.

Bandhakavi, A., Wiratunga, N., Massie, S. & Padmanabhan, D., 2017. Lexicon Generation for Emotion Detection from Text. *IEEE Intelligent Systems.*

Buja, A., Mcdonald, J. A., Michalak, J. & Bellcore, W. S., n.d. *Interactive Data Visualization using Focusing and Linking,* s.l.: s.n.

Burnap, P. et al., 2016. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies.*

By, S., Garg, P. & Bassi, V., 2016. *Sentiment Analysis of Twitter Data using NLTK in Python,* s.l.: s.n.

Conway, B. A., Kenski, K. & Wang, D., 2015. The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. *Journal of Computer-Mediated Communication.*

D3, 2018. *D3 API Reference.* [Online]   
Available at: https://github.com/d3/d3/blob/master/API.md  
[Accessed 9 October 2018].

Diakopoulos, N., Naaman, M. & Kivran-Swaine, F., 2010. *Diamonds in the rough: Social media visual analytics for journalistic inquiry.* s.l., s.n.

Go, A., Bhayani, R. & Huang, L., n.d. *Twitter Sentiment Classification using Distant Supervision,* s.l.: s.n.

Hao, M. et al., 2011. *Visual sentiment analysis on twitter data streams.* s.l., s.n.

Hürlimann, M. et al., 2016. *A Twitter Sentiment Gold Standard for the Brexit Referendum.* s.l., s.n.

Hutto, C. J. & Gilbert, E., 2014. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,* s.l.: s.n.

Java, A., Song, X., Finin, T. & Tseng, B., 2007. *Why We Twitter: Understanding Microblogging Usage and Communities,* s.l.: s.n.

Jordan, M. I. & Mitchell, T. M., 2015. *Machine learning: Trends, perspectives, and prospects.* s.l.:s.n.

Kiritchenko, S., Zhu, X. & Mohammad, S. M., 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research.*

Kouloumpis, E., Wilson, T. & Moore, J., n.d. *Twitter Sentiment Analysis: The Good the Bad and the OMG!,* s.l.: s.n.

Kumar, S., Morstatter, F. & Liu, H., 2013. *Twitter Data Analytics.* s.l.:s.n.

Kwak, H., Lee, C., Park, H. & Moon, S., n.d. *What is Twitter, a Social Network or a News Media?.* s.l.:s.n.

Lansdall-Welfare, T., Dzogang, F. & Cristianini, N., 2017. *Change-Point Analysis of the Public Mood in UK Twitter during the Brexit Referendum.* s.l., s.n.

Liu, K.-L., Li, W.-J. & Guo, M., n.d. *Emoticon Smoothed Language Models for Twitter Sentiment Analysis,* s.l.: s.n.

Nakov, P. et al., n.d. *SemEval-2016 Task 4: Sentiment Analysis in Twitter,* s.l.: s.n.

Palomino, M. et al., 2016. The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to “nature-deficit disorder”. *International Journal of Environmental Research and Public Health.*

Pedersen, S. et al., 2014. *OpenAIR@RGU The Open Access Institutional Repository at Robert Gordon University Citation Details Citation for the version of the work held in 'OpenAIR@RGU',* s.l.: s.n.

Pedersen, S. & Baxter, G., n.d. *Twitter response to televised political debates in Election 2015,* s.l.: s.n.

Pozzi, F. A., Fersini, E., Messina, E. & Liu, B. eds., 2017. *Sentiment analysis in social networks.* Cambridge, MA: Morgan Kaufmann.

Rauch, J., 2016. Are There Still Alternatives? Relationships Between Alternative Media and Mainstream Media in a Converged Environment. *Sociology Compass.*

Severyn, A. & Moschitti, A., n.d. Twitter Sentiment Analysis with Deep Convolutional Neural Networks.

Spencer, J. & Uchyigit, G., 2012. *Sentimentor: Sentiment analysis of twitter data.* s.l., s.n.

Tang, D. et al., n.d. *Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification \*,* s.l.: Association for Computational Linguistics.

Wagh, B., Shinde, P. J. V., Kale, P. P. A. & Student, M. E., n.d. *A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques,* s.l.: s.n.

Wang, C. et al., 2013. SentiView: Sentiment analysis and visualization for internet popular topics. *IEEE Transactions on Human-Machine Systems.*