

Sentiment analysis of social survey data for the City of Casey

M. Lepelaar¹, A. Wahby¹, M. Rossouw¹, L. Nikitin¹, K. Tibble¹, P. Ryan²  and R. Watson² 

¹ Swinburne University of Technology, Hawthorn, Victoria 3122, Australia

² Ryan Watson Consulting Pty Ltd, Vermont, Victoria 3133, Australia

Email: peter@ryanwatsonconsulting.com.au

Abstract: Big Data analytics can be used by smart cities to improve the liveability, health, and wellbeing of their citizens. Smart cities and councils use social surveys and also social media to engage with their communities and these can require sophisticated analysis techniques. The primary direction of this research was focused on using a lexicon-based approach to carry out a sentiment analysis on data from social surveys to produce high-level reports of actionable insights. Data analysis techniques using RStudio were applied to open-source datasets, which included the Casey Next short survey 2016 dataset published by the City of Casey (CoC) and the 2018-20 Social Indicators Survey datasets published by the City of Melbourne (CoM). Initial findings revealed that the qualitative nature of the CoC dataset could produce rich insights using sentiment analysis, whereas the CoM datasets were not suited to this analysis due to their quantitative nature and were not analysed further. These actionable insights could be used by the CoC to compare and validate its own findings and to reveal further insights.

The 2016 Casey data was first cleansed and then RStudio analysis performed to create word cloud visualizations and bar charts for sentiment analysis of the survey questions. The analysis has led to several actionable insights for the CoC. These include that the CoC should: consider safety, cleanliness and family friendliness as their top priorities; invest further in the environment providing more parks and green spaces; and improve transport options for their residents. These findings were then used to inform social media analysis via the Twitter application programming interface (API). The Twitter API was customised to extract tweet data posted by users within a 25 km radius of the centre of CoC's geographical boundaries specific to relevant keywords, to calculate sentiment based on the most recent data.

The R codes were integrated within a Shiny application package to create a set of interactive webapps hosted through the Shiny server which produce (a) word clouds and bar charts of sentiment from the historic survey data and (b) more immediate sentiment analysis as bar charts from Twitter feeds. Due to restrictions and limitations in terms of publishing findings from the Twitter data, the webapps were embedded within a website set up through Hostinger to secure and restrict access only to approved audiences. The main purpose of the Twitter API and website is to provide a web solution that can be customised to estimate current sentiment for key issues for the CoC. The solution can be refined to increase its scalability, accuracy and performance through further development and research.

The project was carried out by a team of students from Swinburne University of Technology for the sponsor Ryan Watson Consulting Pty Ltd (www.ryanwatsonconsulting.com.au). This project was proposed to impart valuable knowledge and skills to the students and prepare them for jobs in the rapidly growing data science market while the sponsor also planned to expand its technical skills repertoire in sentiment analytics to better serve its clients in the smart cities marketplace.

Keywords: *Smart cities, open datasets, data analytics, government, health & wellbeing, Python, RStudio*