

TEXT MINING TWITTER: COLOMBIA 2018 ELECTIONS

JUAN SEBASTIÁN GARCÍA RODRIGUEZ
TRAVIS DUNLOP

Abstract

Between May and June of 2018 the people of Colombia will vote for their next president. As with any modern election, people are using Twitter, the social media platform, to support candidates they like, discredit the others, and debate who should win. Twitter provides a massive open forum to create dialogue across the country. Some groups take advantage of this by creating bots, which automatically post political tweets to in an attempt to sway voters [2]. In our project, we use the text mining skills gained in this course to assess the influence of these bots. We hope to answer two questions: how many bots are there and who are they supporting?

1 INTRODUCTION

In order to answer these two questions we split the project into two sections: bot detection and sentiment analysis. For bot detection we leverage an existing algorithm and use semi-supervised learning to extrapolate it's findings. With sentiment analysis, we hand-labeled a portion of the tweets and use those to help classify the rest into positive, negative, or neutral.

The data we use is a set of [[INSERT NUMBER HERE]] recently collected tweets that mention any of the colombian candidates. By the end of the analysis we report what percent of the tweets support which candidate, and how many likely bots we were able to detect.

2 BOT DETECTION

3 SENTIMENT ANALYSIS

4 CONCLUSION

REFERENCES

- [1] Clayton A. Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. Botornot: A system to evaluate social bots. *CoRR*, abs/1602.00975, 2016.
- [2] Jon Swaine. Twitter admits far more russian bots posted on election than it had disclosed. *The Guardian*, Jan 2018.