Twitter sentiment analysis of the Affordable Care Act in 2017

MICHAEL SMITH[1]

[1]*School of Informatics and Computing, Bloomington, IN 47408, U.S.A.*
Corresponding authors: mls35@iu.edu

The mission of this project is to utilized technologies and cloud computing to perform a successful sentiment analysis through software deployment written in python. The software deployment will encompass data mining, analysis of big data, and comparison of this deployment across a variety of cloud computing services. The sentiment analysis will use the social media platform twitter and python libraries that effectively extract relevant data to the project goal.

Keywords: Cloud, I524

Report: https://github.com/cloudmesh/sp17-i524/tree/master/project/S17-IO-3019/report/report.pdf
Code: https://github.com/cloudmesh/sp17-i524/tree/master/project/S17-IO-3019/code

1. INTRODUCTION

The current political climate of the United States is divided on many important issues. There is a disconnect between the motivations of the politicians of today and what is deemed important to the American people. The affordable care act(ACA) also known as Obamacare has been a target of the GOP, however it is uncertain if that sentiment is shared by most Americans. With a law that affects millions of Americans it is often difficult to gauge how the American people feel about the current healthcare law. Twitter is one of the biggest social medias on the internet with 67 million active users in the United States as of Q4 2016.[1] It is the goal of the project to gauge how Obamacare is viewed by twitter users who reside in the United States.

2. TECHNOLOGIES

Cloudmesh is an open source toolkit that allows the user to work across a variety clouds, virtual machines and clusters. This facilitates the ease of porting deployments to different clouds enabling the capability to benchmark cloud performance on a particular deployment. [2] The project was developed by Gregor von Laszekski and his colleagues at Indiana University. This will be the primary interface used to port the software to clouds such as chameleon cloud and futuresystems.

Ansible is open source software used for automation of provisioning of software deployments. This will be used when deploying on virtualization and cloud environments.

Chameleon cloud and futuresystems to be discussed here.

3. PYTHON

Early scripts have already been developed utilizing Twitters API and the python library tweepy to mine tweets that contain information relevant to Obamacare. Currently, the code authenticates with the twitter API, mines the tweets that contain a keyword of interest and finally a sentiment analysis which will rate a tweet by its polarity and subjectivity. For the sentiment analysis, the TextBlob library is used for its natural language processing(NLP) functionality. The final part of code outputs the tweet content and sentiment analysis into two columns into a csv file. This code will evolve to include data visualization through matplotlib and deeper analysis as the project moves closer to completion. Other python libraries will likely be added as well.

4. BENCHMARKS

Software will be deployed and benchmarked on various clouds.

5. LICENSING

TBD

6. WORK BREAKDOWN

Michael Smith is reponsible for all aspects of this project.

7. CONCLUSION

The software once finalized will be deployed across various clouds with the help of cloudmesh and ansible. Benchmark performance of the various clouds as well as analysis of the

twitter sentiment data will encompass most the final project report.

8. AUTHOR BIOGRAPHIES

Michael Smith is a senior quality control peptide chemist at Creosalus Inc. in Louisville, Kentucky. Michael possesses a MS in pharmaceutical sciences and a BS in Biology from the University of Kentucky. He will obtain his MS in Data Sciences program from Indiana University in May 2018. His current interests are python programming, data analytics, and spending time with his children.

REFERENCES

[1] Statista, "Number of monthly active twitter users in the united states." [Online]. Available: https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/

[2] Cloudmesh, "Cloudmesh," Webpage. [Online]. Available: https://cloudmesh.github.io/