

Analyzing Emotions - Twitter / Reddit / Youtube

Nagalakshmi Prasanna Pujita
Bodapati
Computer Science
Binghamton University
Binghamton, New York, USA
nbodapa1@binghamton.edu

Saisuraj Aitha
Computer Science
Binghamton University
Binghamton, New York, USA
saitha2@binghamton.edu

kshamitha Gandu
Computer Science
Binghamton University
Binghamton, New York, USA
kgandu1@binghamton.edu

Mohith kumar Sopparam
Computer Science
Binghamton University
Binghamton, New York, USA
msoppar1@binghamton.edu

Anshul Upadhyay
Computer Science
Binghamton University
Binghamton, New York, USA
aupadhy5@binghamton.edu

Abstract

The comments sections of social media platform have become the new playground for online bullying. The impact of toxic comments is much more catastrophic than we think. Therefore, having a solid toxicity flagging system in place is important if we want to maintain a civilized environment on social media platforms to effectively facilitate conversations. But, deciding if we have to flag a comment or not is very time taking. If we have an automated process where we can automatically detect abusive keywords in the comments or posts can save the time of website moderators and also it will have a great impact on improving the discussion rather than focusing on using harsh words. So, we are mainly focusing on posts/comments in Reddit, Youtube and Twitter that contains toxic keywords and flag them as negative emotions by taking into consideration of data we have obtained in a period of time.

Keywords: YouTube API, Tweet-Stream, data analysis, Reddit API, GoogleAPIClient, Python, MongoDB, Twitter API, Reddit API

ACM Reference Format:

Nagalakshmi Prasanna Pujita Bodapati, Saisuraj Aitha, kshamitha Gandu, Mohith kumar Sopparam, and Anshul Upadhyay. 2022. Analyzing Emotions - Twitter / Reddit / Youtube. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Recently, people have started using online forums frequently for discussions. Along with the discussions on online forums, trolls and spammers have become more common. It is a time-consuming and tedious job to moderate the comments and posts on these forums, so the organizations have to rely on external people to handle them. We are trying to differentiate between negative and positive comments and tweets on Twitter, Reddit, and YouTube. To achieve this, we would get real-world data from the above mentioned platforms. We also intend to visualize Twitter, Reddit, and YouTube's data.

2 METHODOLOGY

We used SampleStream to gather tweets from twitter, GoogleAPIClient to extract random comments for a particular video and used direct Reddit oauth authentication and directly hit the API to collect data on subreddits by generating a query and collected the posts from these subreddits. Once, the complete data is collected we stored in the MongoDB by creating separate database and collection for each datasources we have chosen.

We have collected an average of 6 million tweets and approximately 2 million YouTube comments and around 1 million reddit posts over the past seven days. The data will be then extracted from MongoDB and cleaned by implementing a regex function. By using this function, we will now remove the noise from the data collected. For example, this function can be used to remove data that is not in english language. Now we use this updated clean data to perform the analysis to find emotions whether a particular tweet/comment/post is positive, Negative or neutral. Once the complete data analysis is complete we will represent the data collected in a graph.

3 CONCLUSION

So, finally the main aim is to perform data analysis on the extracted data from Reddit, Youtube and Twitter. Using various tools we will represent the results and outcomes graphically.