



## INTRODUCTION

Hashtags have become a common way of aggregating a group of social media users around a common topic. Most social media sites have incorporated this feature and it is now starting to be used by marketing agencies as way to engage with its audience and gauge their response to their campaigns. I wish to make this task simpler by developing a ranking system with aggregates data from multiple social media outlets to deliver a consistent view of the engagement both quantitatively (like the number of people using a particular hashtag, etc.) and qualitatively (like classification of the use as a positive or a negative response).

## OBJECTIVE

The primary goal of the project is to develop a ranking system based on sentiment analysis to provide an indicator of the engagement of a hash tag based marketing campaign with its intended audience.

## METHODOLOGY

The main steps to achieve the desired objective are outlined as follows:

- Setup the developing environment to capture social media posts containing a hashtag.
- Preprocess the text captured to perform sentiment analysis.
- Identify and learn tools to perform sentiment analysis to extract data regarding the users emotions and polarity towards the particular product represented by the hashtag.
- Identify tools and techniques to develop a ranking system to determine the impact of the hashtag on its users based on the emotion and polarity data extracted in the previous step.

## TOOLS AND TECHNIQUES

The following tools were used to develop the project:

- R as the base language
- Packages “twitterR” and “Rfacebook” for capturing social media data in R
- Package “sentiment” for performing classification of individual tweets based on emotions and polarity.
- Packages “wordcloud”, “ggplot2” and “RColorBrewer” for graphically representing the results.

## PROGRESS AND RESULTS

Using the aforementioned tools I was able to perform basic emotional and polarity classification on a corpus of 100,000 tweets related to ‘#TheFlash’ hashtag collected on the night the season finale of the TV show The Flash was aired. The R package Sentiment commands `classify_emotion` and `classify_polarity`; which use a naive Bayes classifier trained using Carlo Strapparava and Alessandro Valitutti’s emotions lexicon and Janyce Wiebe’s subjectivity lexicon; were used to classify the emotions and polarity of the given corpus. The results of the analysis are presented below.

Table 1: The Results of the classifier scores and best fit emotions

Tweet No.	ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE	BEST_FIT
486	1.468718	3.09234	2.067836	1.025478	1.727707	7.340836	surprise
634	1.468718	3.09234	2.067836	1.025478	7.340836	2.786959	sadness
646	1.468718	3.09234	2.067836	13.65619	1.727707	2.786959	joy
775	7.340836	3.09234	2.067836	1.025478	1.727707	2.786959	anger
930	1.468718	3.09234	2.067836	1.025478	7.340836	2.786959	sadness
1217	1.468718	3.09234	7.340836	1.025478	1.727707	2.786959	fear

Table 2: The Results of the classifier scores and best fit polarity

Tweet No.	POS	NEG	POS/NEG	BEST_FIT
486	24.97756	9.47547	2.636023	positive
634	1.031278	26.84236	0.03842	negative
646	24.97756	0.445453	56.07224	positive
775	17.22652	9.47547	1.818012	neutral
930	8.782323	26.84236	0.327182	negative
1217	8.782323	17.81234	0.493047	negative

Table 3: The Processed Tweets and the classified emotion and polarity

Tweet No.	The Tweet	Emotion	Polarity
486	theflash season finale was surprising didnt see all that coming and the golden age helmet emotional and visual awesome guys	surprise	positive
634	theflash ep on heartbreaking finale death seasons speedsters and villains	sadness	negative
646	that was a great season finale so so good theflash	joy	positive
775	season finale of theflash was just high drama of events paradox ending	anger	neutral
930	theflash season finale was full of heartbreak death and a whole lot of confusing science	sadness	negative
1217	theflash pulled a walkingdead and made us dread this summer because new episodes come back in the fall	fear	negative

The bar plots obtained for the results of the complete set of outputs is presented below in fig. 1.

## FUTURE WORK

The future direction is to develop a relative ranking system or impact factor calculator based on the data analysis performed till now.

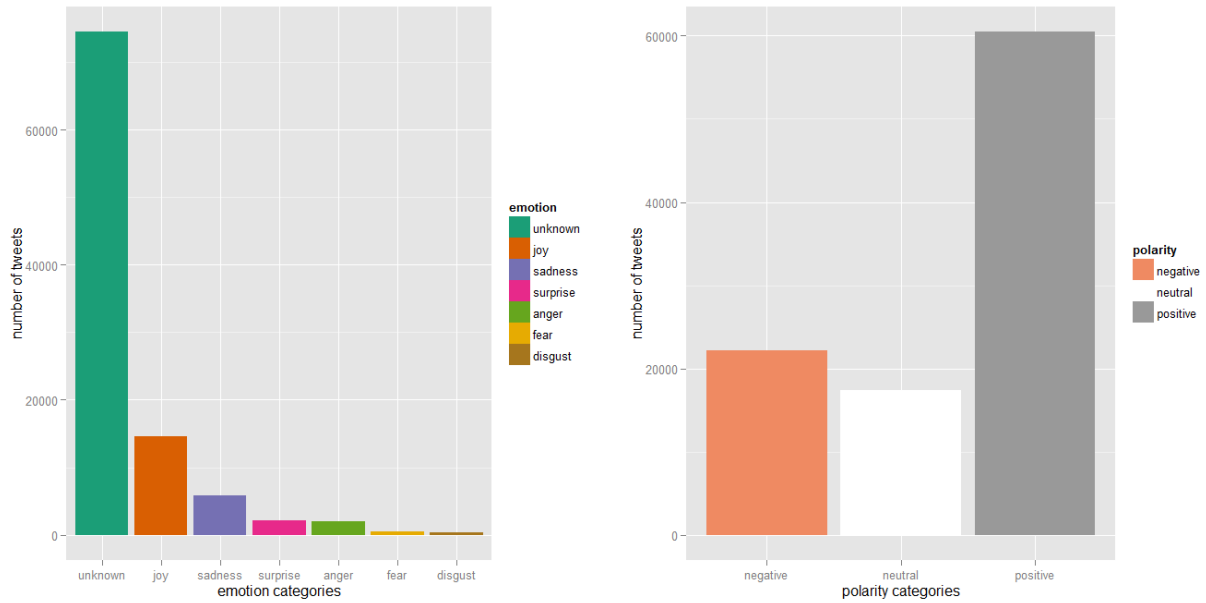


Figure 1 The Bar graphs showing the final polarity and emotional classification of the collected tweets.

## BIBLIOGRAPHY

1. Sanchez Gaston, Mining twitter with R, <https://sites.google.com/site/miningtwtter/home>
2. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
3. Jeff Gentry (2015). twitteR: R Based Twitter Client. R package version 1.1.8. <http://CRAN.R-project.org/package=twitteR>
4. Timothy P. Jurka (2012). sentiment: Tools for Sentiment Analysis. R package version 0.2. <http://CRAN.R-project.org/package=sentiment>
5. H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.