

Homework # 5

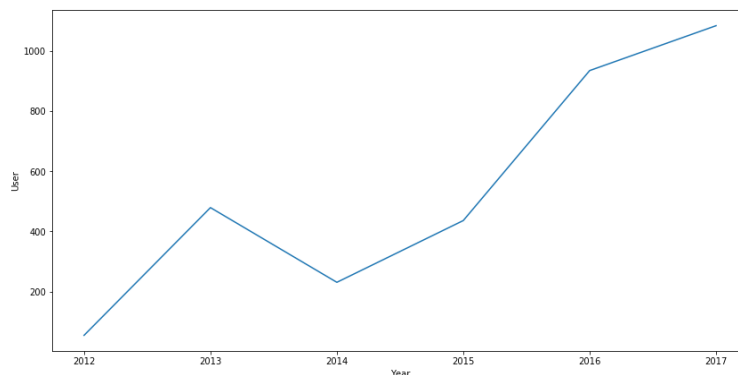
Introduction

Social media platforms such as Twitter, generate large volumes amount of data and an extremely high rate based on the events that take place around the world. As you are likely aware, organizations are constantly trying to find ways to analyze the information that is produced by social media users in support of business objectives. Beyond trying to get a pulse of current events and the sentiment felt by their customers, organizations also look to Twitter to gauge how the public feels about emerging trends and opportunities.

One such topic of interest at this time is the use of Artificial Intelligence (AI). The promise and capabilities of the technology is easily matched by the challenges and fears associated with concerns related to privacy, security, governace and ethics. There are a number of individuals and organizations that seek to harness the power of AI to advance their bottom lines, but they are also extremely conscience of the negative perceptions of AI when it comes to implementing its potentially “invasive” approach to procuess and user experience improvement.

Elon Musk is one of those invididuals who’s been know to be on the forefront of emerging technologies, including AI, and not afraid to share his opion on such matters. By analyzing the Tweets he has produced on the topic of AI and leveraging sentiment analysis techniques to determine their sentiment, we can achieve the following objectives:

- Gain a sense of Elon Musk’s perception of AI and its usage in the real world
- Evaluate the ability for technology to interpret the sentiment of short messages, such as Tweets;
- Identify approaches to supplement technology with human intellect to achieve complex teasks



Elon Musk tweets between 2012 and 2017

To address this last point, Amazon’s Mechanical Turk platform was leveraged to assess the viability of leveraging crowdsourcing as a supplemental mechanism.

Analysis

About the Data

The Elon Musk tweet data was originally sourced from [Kaggle](#) and contained more than 3200 tweets with the following attributes:

- **Row ID:** The unique (sequential) identifier for the row
- **Tweet:** The textual content of the tweet
- **Time:** The date and time the tweet was posted to the Twitter platform
- **Retweet from:** The name of the Twitter user, Elon Musk's tweet was sourced (retweeted) from
- **User:** The user name who created the Tweet. Elon Musk

For our purposes, the attribute of greatest importance was "Tweet" and while reviewing the data, we noticed the following key considerations:

- Tweets often contained hash tags (#) and user handles (@)
- Tweets frequently contained URL links to reference additional information
- Postings may have one or more line breaks embedded in the tweet
- Tweets could contain emojis / special characters that are not interpretable or cannot be directly translated to English

Cleaning and Prep

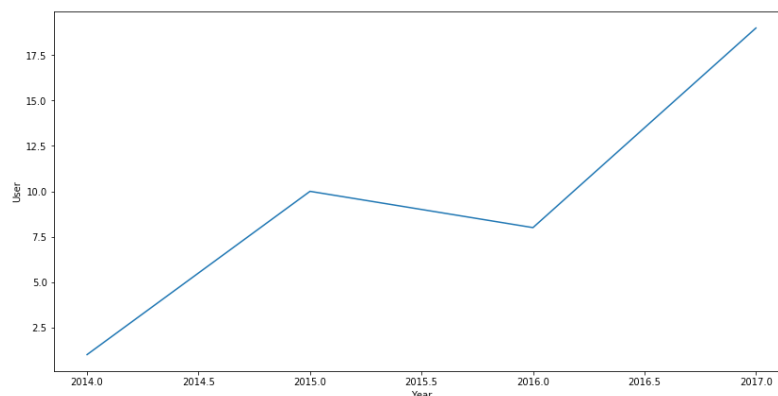
Cleansing

The following two steps provide all of the initial data cleansing required for processing:

- All attributes except for "Tweet" were removed from the data set.
- Special characters and line breaks were removed from the resulting tweets.

Filtering

As previously mentioned, our focus is on artificial intelligence and therefore the tweets were filtered to only include those posts that made mention of the topic. This was performed by searching for tweets that contained either the text string " AI " or "artificial intelligence". After filtering for these two exact phrases, the data was reduced to 38 tweets.



Elon Musk tweets related to artificial intelligence between 2012 and 2017

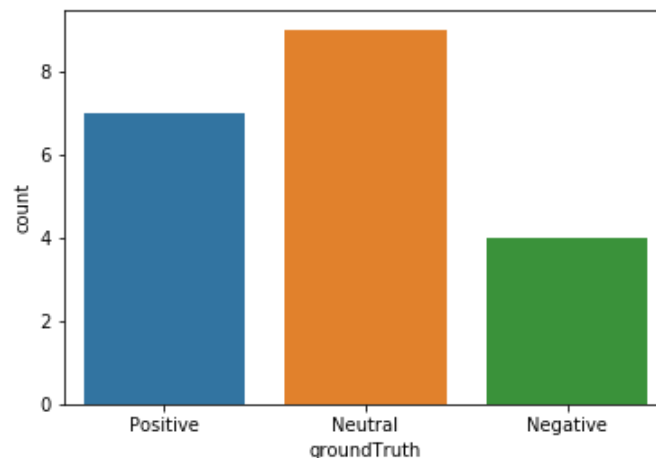
Models

To determine the viability of the use of Mechanical Turk platform, we utilized the following approaches.

Manual / Ground Truth

Each of the 38 tweets from the Elon Musk AI tweet data set was manually reviewed by a single individual to produce the ground truth values by leveraging the following guidelines:

- Statements that conveyed facts (especially related to retweets) were considered neutral.
- Special attention was paid to sarcasm and the use of negation words in the Tweets
- Disagreement with other tweeters leveraging the @handle were considered to be negative (intensity was not measured)
- Recommendations and promotions were considered to be positive, when the tweet was not considered to be simply raising awareness.



Mechanical Turk Workers

During the creation of the Mechanical Turk assignment, the following parameters were used:

- Only the first 20 Elon Musk AI Tweets from the data set were submitted
- Each Tweet was considered an individual task and with 4 workers requested to evaluate each
- Workers were requested to determine if the Tweet represented a positive, negative or neutral sentiment
- Workers were not required to have any specific experience while performing the labeling
- Each task was estimated to take 2 – 3 minutes
- Each completed and accepted task paid \$.25

VADER and SentiStrength Models

The VADER and SentiStrength models were retained from our previous analysis to serve as additional data points to compare the Mechanical Turk results against.

Vader will produce four values for each text analyzed: negative, neutral, positive and compound.

- **Negative:** Compiles a final score for all tokens with negative sentiment

- **Neutral:** Compiles a final score for all tokens with neutral sentiment
- **Positive:** Compiles a final score for all tokens with positive sentiment
- **Compound:** Summarizes the negative, neutral and positive sentiment scores into a single value; and then the value is scaled between -1 and 1.

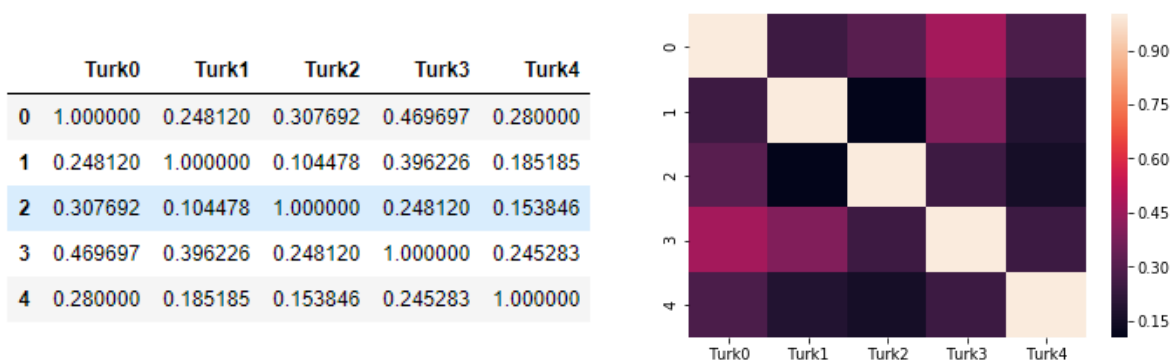
SentiStrength provides values for both positive and negative sentiment contending that the same text can consist of both positive and negative sentiment simultaneously.

- **Negative:** The value, which ranges from -1 to -5 conveys the strength of the negative sentiment identified within the social media text.
- **Positive:** The value, which ranges from 1 to 5 conveys the strength of the positive sentiment identified within the social media text.

Results

Mechanical Turk Workers

The task took less than 20 minutes to complete with varying levels of agreement across the workers. Utilizing the Cohen Kappa score to evaluate the agreement amongst Turkers is represented in the matrix below.



We notice that Turker-3 and Turker-0 demonstrate the greatest amount of agreement amongst any of the workers, while Turkker-1 and Turker-2 demonstrate the greatest disagreement. This chart highlights how the opinions of crowd sourced workers can vary greatly when considering a task as a whole.

Ground Truth

When comparing the results from the Mechanical Turk workers against ground truth we witness the following results.

Worker	Kappa Value
Turker0	0.3023255813953487
Turker1	0.34545454545454535
Turker2	0.3129770992366413
Turker3	0.17602996254681658
Turker4	0.607843137254902
Average Kappa Score	0.03489260651776508

Unlike the level of agreement that was witnessed between workers 3 and 0, the ground truth is most aligned with the Turker 4. This provides yet another example of how the results from different individuals can vary greatly while working with text analysis.

Conclusion

For example, when we look at a comparison of the most frequently chosen answer from the Mechanical Turk works with the ground truth label, we see an increased kappa value (0.5833333333333333) when compared to any individual worker. The Mechanical Turk platform provides a simple means through which tasks that may require human intervention to produce more accurate results than solely relying on analytical models and algorithms.

Mechanical Turk provides a viable option for leveraging crowd sourced resources to perform manual tasks and its configurability allows for increases in throughput with the potential sacrifice of quality. Specifically considering the use case of Tweet sentiment and the likely volume of tasks need manual classification, the use of any worker provides an economical option.

Needless to say, to further explore the benefits of Mechanical Turk it would be helpful to understand the speed and Kappa scores associated with workers with specific experiences and compare the results against the generic pool of workers used within this analysis.