Diego Valdes

IST 736

Home Work 5

12/7/2019

**Introduction**

If one were to consider for a moment, the amount of data they generate in a day, they may be surprised at the footprint they are leaving without even knowing. Consider, for a moment, all of the internet searches a person does in a day for shopping, dining, and random questions. Then add the social media posts. Top that off with any pictures taken that day. Lastly, lets add any third party indirectly collecting data, like traffic lights as an example. Repeat that for everyone in the developed world, and suddenly that’s a whole lot of information to comb through. Now imagine, if through all of that data, a company was tasked with finding a particular piece of information to assist them is improving sales. The analogy ‘needle in a haystack’ may seem appropriate.

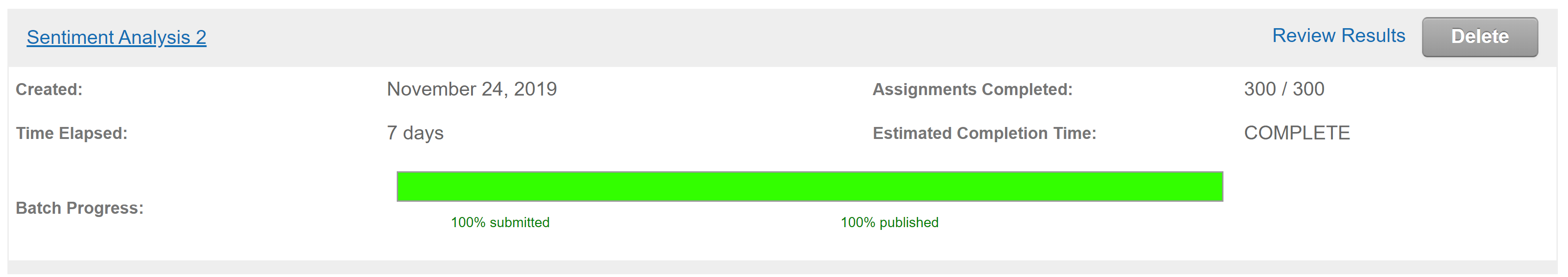
That is where artificial intelligence comes into play and why it’s one of the buzz words in every industry today. It is a possible solution to problems that didn’t exist in the past, as well as for situations where humans are simply not the right tool to accomplish the task. In most of these situations, humans are the best suited, but the volume of the work and the time constraints make it impossible to rely on humans. Think of the man power required to read all the tweets posted in one day to search for something inappropriate; there may not be enough humans on the planet to accomplish that task in a timely manner.

However, even though AI is becoming more important in the lives of humans, there are tasks that the computer simply can’t do. While it can ‘learn’ to make predictions, that learning is based on the feeding of data that has a series of attributes and an outcome. A computer can’t generate the outcome of data it has never learned to analyze. That is something humans can do. While a machine needs examples to be able to perform sentiment analysis, a human need only read a few lines to make that determination. While humans are limited by the amount of time it may take them to do tasks, there are simply some tasks that a machine cannot be relied upon to complete.

**Analysis**

**About the Data**

The data used a cleaned version of movie reviews obtained for previous analysis and submitted to Amazon Mechanical Turk for human sentiment analysis. The file was stripped of all predictions and only 100 of the 2000 reviews were submitted for human analysis. This was due to the cost of providing 5 reviewers for 2000 workers - $120. Instead 3 reviewers for these 100 reviews was submitted at a cost of $6. Each reviewer provided a sentiment of Neutral, Positive, or Negative. The 100 reviews submitted were all originally ranked negative.



A visual examination of the Turk file found some positive sentiments. Although no proof can be provided and due to the speed of the analysis being completed, it may be probable that this is due to many of the workers guessing at the sentiment. The Turk file was also reformatted as it was provided at the grain of a workers analysis and for comparison of results the grain needed to be on the review level.

**Models**

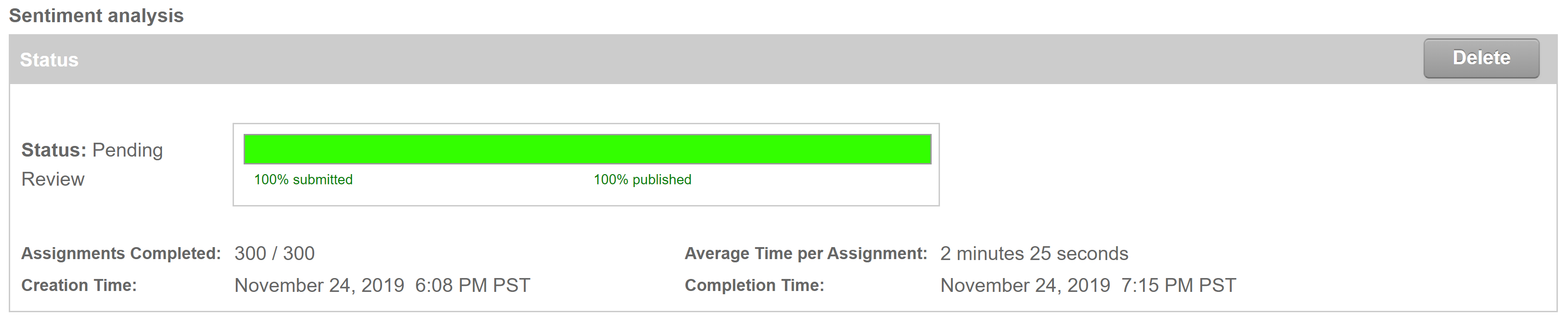
A Cohen Kappa Score was run on comparing the Turk workers with their peers and the original sentiment, which was always negative.

**Results**

There wasn’t much consistency between any comparison. This result supports the initial assumption that perhaps the sentiment provided was the result of the workers guessing.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Worker 1** | **Worker 2** | **Worker 3** | **Original** |
| **Worker 1** |  | 0 | -2.22 | 0 |
| **Worker 2** | 0 |  | 0.01 | 0 |
| **Worker 3** | -2.22 | 0.01 |  | 0 |
| **Original** | 0 | 0 | 0 |  |

Further examination of the results provided by Turk on the average time to complete the assignment was within the realm of reason to rule out any ‘guessing’ on the part o the workers. This, however, does little to explain the disagreement between the workers and original results.



**Conclusion**

Amazon Mechanical Turk is a good tool to get a quick and large workforce to do tasks that a computer cannot do. The service also provides ways to limit the types of workers that can pick up a project, thus allowing the requester to better have control over the pool of workers going over their data.

While this is a great tool for a data professional to have, it does leave some questions to be answered. Having the workers deviate greatly from the original is of some concern. Before deciding if Turk is a reliable go to for this kind of work, it would be best to determine why this result was attained. Had the workers kappa score been closer to each other, this may be a moot point, as neutral was an option for the analysis for Turk but not one in the original data.

In the end, the final say on Amazon Mechanical Turk is that it is a tool worth knowing and exploring, but until definitive answers can be determined on why the inconstancy of the results were obtained here, it is advisable for data professionals to tread carefully in taking the results from it as gospel.