Rohini Shrivastava

IST 736

**Introduction:**

Artificial Intelligence is a branch of computer science that is interested in creating machines that can accomplish tasks done by humans and animals. Alan Turing is credited with the idea of the concept when he created the idea about a “thinking machine” that could reason like a human can. In the 1950’s, he created the Turing Test that is used to help determine whether a machine’s behavior is indistinguishable from a human. Machine learning is a subsection of AI. It is an application that uses computer algorithms to learn and improve from it’s experience without being explicitly programmed. Artificial Intelligence is a very controversial topic regarding how it can help automate tasks, but also the ethics behind it.

Artificial Intelligence has multiple positives associated with it. Machines can overcome the limitations humans have; the only thing needed is enough power to keep them going. They can do repetitive tasks that can reduce labor costs. They can also help humans in everyday items. There are many applications, such as GPS, Google Home, Alexa, etc., that can help us in directions, reminding on tasks and answering simple questions. As machines have no emotions, they can make decisions based on certain criteria instead of emotions. This is seen used in healthcare when diagnosing patients with illnesses.

Along with the positives we also have multiple negatives. As machines have no limitations and can reduce labor costs it can cause higher rates of unemployment. People around multiple companies are trying to automate their tasks, which can cause people doing those tasks to lose their job. One such industry that has seen this are call centers. There are now automated messages and chat bots that can answer many questions, reducing the need for human interaction. Many places have laid off workers as the need for them no longer exist. Machines also show no creativity as they can only do what they are taught. As they have no emotions, they cannot create ideas and show critical thinking in situations. They cannot tell what is ethical or legal and have no judgmental skills (*West, 2019)*

So what kind of thoughts do everyday people have about Artificial Intelligence? With movies, games, and tv shows, there have been many different thoughts on what AI is. To find out how the general public feels about AI we can use various Sentiment Analysis tools.

**Analysis:**

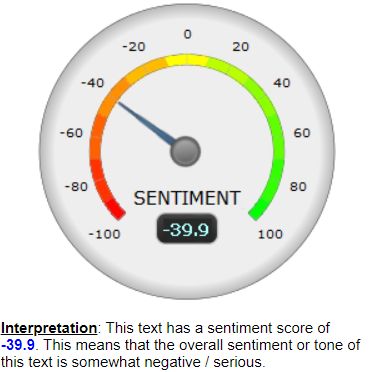
About the Data:

To understand the how people feel about Artificial Intelligence tweets were gathered from Twitter. There was a search done for tweets that mentioned “Artificial Intelligence” and were written in English. The tweets taken need to have “Artificial Intelligence” or “AI” directly in the tweet and not in a hashtag. After finding such tweets, the first 50 were exported the into a word document. From there the tweets had removed any emoji or punctuation, such as a period, comma, semicolon, hashtag, etc. If the tweet contained a link or a picture, those were removed as well.

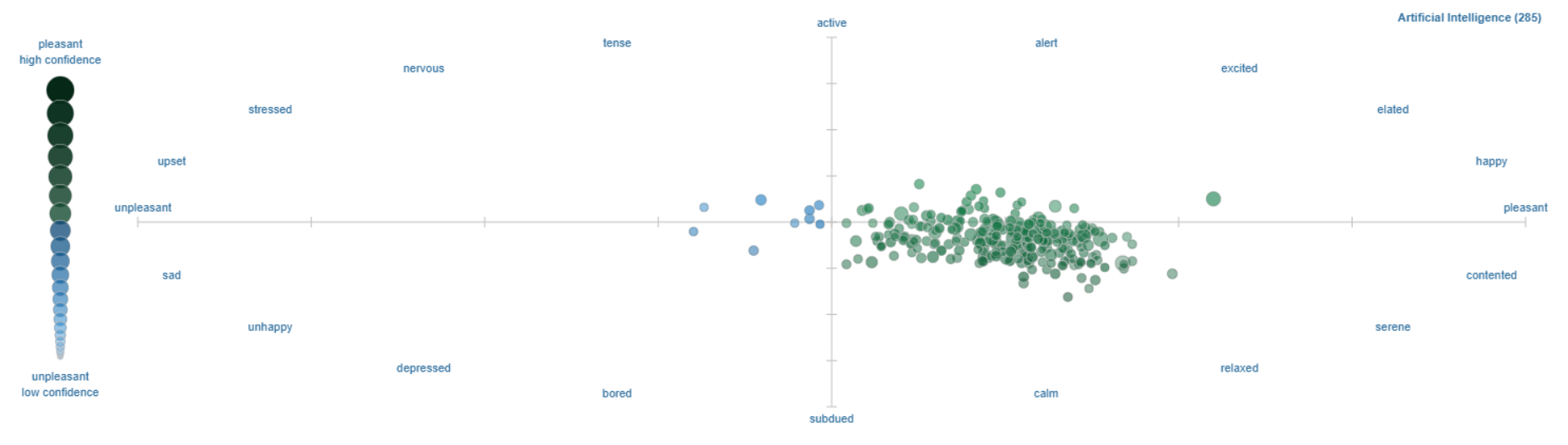
Models:

Sentiment Analysis Tools:

*DanielSoper Sentiment Analyzer-* The document of tweets were copied and pasted into the Sentiment Analyzer. This analyzer looks at a text and scales it between -100 and +100. Having a score of -100 shows a negative/serious tone about the text, whereas +100 is positive/enthusiastic. After putting the tweets about Artificial Intelligence into the tool, the text had an overall sentiment of -39.9.



*Figure 1: Sentiment Analyzer for collected tweets*

*Tweet Sentiment Visualization-* The Tweet Sentiment Visualization (Sentiment Viz) looks at recent tweets about a certain topic on Twitter and analyzes them. In this case, the tweets were looking at Artificial Intelligence. This API takes the tweets and sorts them by pleasant/unpleasant and places them in different categories. Most of the tweets seem to be pleasant, however they were more subdued than active. 

*Figure 2 Sentiment Viz for Recent Twitter Tweets*

SKLearn/CountVectorization and LDA:

*SKlearn* is a python programming library that is used for classification, regression, and clustering algorithms. It uses supervised and unsupervised algorithms. When doing Text/document classification, sklearn uses supervised machine learning to assign categories to different documents. Inside of sklearn there is a module called CountVectorizer. *CountVectorizer* converts a collection of documents into a matrix of tokens and their counts. In this module the documents can be passed as a filename, file or content. The default is content such as a string or byte. If a filename is passed, it is expected to be a list of files so it can fetch the raw content to analyze them. If a file is passed, the items have to use a read method to put the content of the file into memory. Having the words vectorized gives an accurate representation of what the document contains (*Shaikh, 2017)*.

*Latent Dirichlet Allocation (LDA)* is a type of topic model that is used to classify the text inside a document towards a certain topic. Each topic represents a set of words and LDA just takes them and maps them. LDA can be used to recommend books based off past reading and cluster and understand similar articles *(Ganegedara,2017)*.

**Results:**

The Sentiment Analyzer gave a score of -39.9 on 50 collected tweets. This shows that most of the tweets collected were showing a negative light on Artificial Intelligence. However, it was not a strong negative sentiment as it seems it fell closer to neutral than negative. When the Sentiment Viz was ran, the289 tweets used were shown to have a more positive sentiment based off the color. The pleasant tweets are seen in green and the negative ones are shown in blue. Looking at Figure 2, all the green (pleasant) tweets were clustered together and the blue (negative) tweets were seen as an outlier. As this model gave more information, the tweets were also able to be put in categories. Most of the tweets were subdued rather than active. They fell mostly towards the calm category.

**Conclusions:**

Overall, there are mixed results for the tweets regarding Artificial Intelligence in the Sentiment Analyzer and Sentiment Viz. The tweets that were collected for the Sentiment Analyzer showed more of a negative sentiment, whereas the tweets taken directly from twitter without changing anything for Sentiment Viz showed a more pleasant sentiment. As there were two different sets used, there can be variability. Though the two sets showed different sentiment, they both fell into the serious/subdued category so the tone between the sets were similar.

There is a difference in the sentiment between the two sets. The data set collected manually for the Sentiment Analyzer looked for a specific type of tweet, while the Sentiment Viz collected all the most recent tweets. There also was a difference in the size of the data set. The set used in Sentiment Viz was about five times larger than the Sentiment Analyzer set.

To analyze future sets, it would be best to get a large amount of data to fully encompass the thoughts people have around AI. Many of the tweets seen contained external links or pictures about their thoughts, which can be hard to determine the sentiment for with the tools currently used. In the future, utilizing the sklearn library and vectorizing the data set of tweets can show a general feel for how users think about AI and see the most commonly used words. From there we can also use LDA to categorize the tweets and see how they are related.

**Bibliography:**

Pros and Cons of Artificial Intelligence - A Threat or a Blessing? (2019, September 25). Retrieved January 23, 2021, from <https://data-flair.training/blogs/artificial-intelligence-advantages-disadvantages/>

West, D. (2019, October 25). What is artificial intelligence? Retrieved January 23, 2021, from <https://www.brookings.edu/research/what-is-artificial-intelligence/>

Shaikh, J. (2017, October 30). Machine Learning, NLP: Text Classification using scikit-learn, python and NLTK. Retrieved January 23, 2021, from https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a

Ganegedara, T. (2020, December 05). Intuitive Guide to Latent Dirichlet Allocation. Retrieved January 23, 2021, from https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158