**ML-Assisted Knowledge Extraction of Events and Public Reception.**

**1. Abstract**

In the age of social media, Twitter is a medium for the everyday person to not only voice their opinion on large topics, but also connect with and influence thousands as a result. In 2020, a year of elections, lockdowns, and vast differences of opinion, Twitter was ranked 4th in the world for engagement as external influence through social media resulted in historical events that weren’t all positive. In this research we utilize Machine Learning and Data Science techniques to analyze tweets throughout the year and visualize their correlation with large events and public reception. The project emphasizes the use of algorithms for the purpose of finding what influences people; whether that is classifying harmful bots that spread misinformation or analyzing text sentiment to quantify the attitude of users. The research is currently in progress, but preliminary results suggest a strong polarization of opinions throughout the year, as well as an increase in misinformation spreading twitter bots. Our goal is to use this research to warn about the harmful influence of social media in polarizing people’s opinions on political topics.

**2.1 Data Collection**

The primary initial problem to solve is how to access Twitter data from the past year. This is hindered especially by Twitter’s recent blocking of all mining ability for tweets older than a week, creating a significant initial blockage for the study. Furthermore, the changing of web scraping endpoints in November 2020 disabled a large proportion of regularly used Twitter Web Scrapers. One scraper, however Twint, was fixed using endpoint workarounds by the general Data Science community. The Twint API makes it possible to not only mine any Tweets using given code words, but also to have a continuous scrape without rate limitations. As such, over 5 million tweets regarding two separate events were downloaded to get an initial view of the Twitter data on two large events.

Events were chosen and analyzed based on expected impact on the community. For example, events include the Coronavirus Briefing by President Trump on March 13th, 2020 and the U.S. Election on November 11th, 2020.

**2.2 Machine Learning Analysis**

The tweets regarding all events were collected from before and after the expected “climax” of tweets by as much as 24 hours in order to establish a control for the important trends. Trends looked for were regarding the ability of the event to influence other uses, including mapping the most frequently used terms, calculating a data analysis