Team: **Abnormal Distribution**

Members: *Charles Phelan, Krish Shah-Nathwani, Jacob van Steyn, Matthew Reynolds*

**Why this project? (who benefits & what’s the need):**

The goal of this data science term project is to perform sentiment analysis on tweets from powerful individuals such as celebrities, influential figures, the wealthy, and other notable figures. By analyzing the language used in their tweets, we want to try to determine their political leanings and categorize them as leaning towards the left or the right. This project is an opportunity to apply natural language processing techniques and gain insights into the political sentiments of people in positions of power. Additionally, it may also shed light on the role social media plays in shaping public opinion.

**What data (or datasets)?**

We will be scraping twitter data from many notable public figures. We will obtain as many tweets as we can for each account and use them later in our analysis.

**What is your "data science" toolkit? You should list specific tools / packages you will use.**

* Python
* Tweepy (Python twitter package)
* Matplotlib
* Numpy
* Prettytable
* Other packages as needed…
* Tensorflow or Pytorch (for inference)
* Huggingface Pre-trained models

**Preliminary sketch of what you hope to build:**

Our general outline of what we want to build is as follows:

1. Find people we want to collect tweets from
2. Scrape & store the tweets
3. Find a generally trained sentiment analysis ML model
4. Re-train the model on political tweets from known political figures (\*people who identify as left/right leaning)
5. Analyze collected tweets to classify each individual account
6. (Time Permitting) Classify individual tweets

Elevator Pitch (To be workshopped before presenting):

Hi everyone, my name is Charles and I am here with my team members Krish, Jacob and Matthew. Today, we would like to present our data science project which aims to perform sentiment analysis on tweets from powerful individuals such as celebrities, influential figures, the wealthy, and other notable figures.

The objective of this project is to determine the political leanings of these individuals by analyzing the language used in their tweets. To achieve this, we will be utilizing the Twitter API and Python to gather data and store it. Additionally, we will be using tools such as Matplotlib, Numpy and Prettytable to analyze and visualize the data.

For the sentiment analysis, we will find a pre-trained sentiment analysis machine learning model, which we will then fine-tune using a dataset of political tweets from known political figures. This will enable us to re-train the model to classify the accounts of our chosen notable figures based on their predicted political leanings.

In conclusion, this project will provide us with an opportunity to apply natural language processing techniques, gain insights into the political sentiments of people in positions of power, and explore the role of social media in shaping public opinion. We are excited to share the results of our work and hope to provide valuable insights through this project.

* **Intro -- (Charles)**
* Presentation by Charles and team members Krish, Jacob, and Matthew.
* Aim: Perform sentiment analysis on tweets from powerful individuals.
* **Data collection -- (Matthew)**
* Utilizing the Twitter API and Python to gather data. Would like to collect ~100 tweets per selected account, ~10k total assuming 100 selected accounts.
* Store data for analysis in mongodb on a per-account basis.
* **Data Analysis -- (Jacob)**
* Using tools such as Matplotlib, Numpy, and Prettytable to analyze and visualize the data.
* Will find a pre-trained sentiment analysis ML model, then re-train it using a dataset of political tweets from known political figures.
* Will use the model to classify notable figures' accounts based on their predicted political leanings.
* **Conclusion -- (Krish)**
* Excited to share the results of the project.
* Provide valuable insights into the political sentiments of people in positions of power.
* Explore the role of social media in shaping public opinion. (make sure to say this)
* Opportunity to apply natural language processing techniques.