American Medical Association (AMA)

The American Medical Association (AMA) is a professional association and lobbying group of physicians and medical students. The Core Purpose of the AMA is, "To promote the science and art of medicine and the betterment of public health." (https://www.ama-assn.org/)

Project Description

The HDAG team will use data collected from Twitters Historical Powertrack API programmatically to classify Twitter users as Physicians, Residents & Fellows, Med Students, Health Care Providers/Other Clinicians (Medical Coders, Nurses, PA's,Etc.). This could be accomplished using a Twitter users Bio and a sample of their most recent Tweets. AMA currently has a semi-robust method to identify Physicians and their specialty using string matching coupled with a filtering methodology but would like to improve upon this process and automate data ingestion/classification. Our team would build on this to remove junk contributors (users creating content for contents sake) and identify key influencers using data metrics. Our team will classify an existing list of Twitter users and develop a data pipeline to ingest new Twitter users and classify them as well.

Internal Partners: Data Science Project Manager of Strategic Insights

Datasets: AMA-provided dataset of Tweets (unclassified and classified)

Coding Languages: R (w/ Tidyverse) (required), Python, experience w AWS (optional), SQL (optional)

Specific Skills

1. Machine Learning: Using machine learning algorithms to classify data.
2. Natural Language Processing: analyze and process language for ML

Expected Technical Difficulty: **Intermediate**