INSIKT INTELLIGENCE

Leveraging Data for a Safer World
The most accurate AI Analysis on the market

CASE STUDIES

ASIAN-AMERICAN HATE

Natural Language Processing

Network Analysis

Deep Learning

In the recent months, the Asian-American community in in the United States has seen a surge in online harassment, physical harassment and discrimination, sometimes correlated with the unfolding of events related to the COVID-19 pandemic. Not only has hate speech against Asian Americans been running high on social media and other parts of the Internet, but it has also been shown to be a precursor and, at times, a catalyst for physical attacks.

PROBLEM

Confronted with this situation, a civil rights association contacted us to provide some insights as to whether **the animosity was originating in specific identifiable online circles** and if the resentment of the most crude messages was constant or pivoted around specific events (e.g. if there were peaks in their frequency).

HOW WE HELPED

With this brief, we focused on 2 major areas: on the one hand, through our **hate speech model**, detecting in which instances hate discourses were present and, within those, how pervasive hate speech was, and on the other hand, analyzing if there were any **clusters of users** that were particularly active in hate speech production.

Through data and text mining, our model identified specific despective keywords that were applied to Asian Americans (“china virus”, “chonky”, “lingling” “Kung flu”, “ching chong”) and we monitored their evolution and frequency on social media (e.g. Twitter, Gab, Stormfront, 4chan, Bitchute, 8kun) checking to see if there was a **correlation between spikes in their usage** and the time and day of news reports covering **real life events**.

> Seemingly benign practices, such as the use of emojis and GIFs, were also being used to (re)produce hate.

Diagnosing what specific types of events trigger surges in online hate volumes provides an **enhanced situational awareness** that can help better inform what strategies should be adopted to monitor these detected trends and, ultimately, address them.

These insights proved to be even more helpful when through **pattern identification** we were able to determine that seemingly benign practices, such as the use of emoji and GIFs, were also being used to (re)produce hate.

SOLUTION

By conducting social network analysis, we detected that a high number of **seemingly unrelated Youtube accounts were actually part of a covert network of reactionary racist influencers, which showed up as a cluster of users** in our network visualization tools. Unfortunately, Youtube was not the only platform where hate speech is present, and coordinated harassment has also been **pervasive on Twitter**.