**Thesis:** Monitoring Hate Crime Against Asians During the COVID-19 on Twitter

**Abstract:** The COVID-19 started from Wuhan, China, in late 2019, and after being utterly contagious in Asian countries, it rapidly spread to other countries. This disease caused governments worldwide to declare public health crisis with drastic measures taken to contain the spread of the disease. This pandemic affected the lives of millions of people. Many citizens that have lost their lives and jobs are going through a wide range of emotions, such as disbelief, shock, concerns about health, fear about food supplies, anxiety, panic, etc. All of the aforementioned new incidents and phenomena led to the spread of racism and hate against Asians in western countries, especially in the United States. The statistics show that Anti-Asian hate crime in 16 of America's largest cities increased 149% in 2020.

We present an approach to measure and monitor Americans' hate crimes against Asians on Twitter. We have downloaded 10 million tweets by Twitter API V-2, and we hand-labeled 3000 tweets to train our system. As a baseline for our research, we have adopted C. Ziems's work, and to achieve an accurate training dataset, we relabeled their annotated dataset by our four Asian annotators. Then we found the agreement rate for each tweet between our four annotators and Ziems's labels. We also calculated the Fleiss'Kappa number to determine the total agreement between annotators in our labeled dataset, which is 0.36 and is interpreted as fair. We improved this number by finding thresholds of the agreement for biased classes to remove tweets from those classes. Although the problem of biased datasets on abusive language detection has been addressed more frequently, biases arising from trained classifiers have not yet been a matter of concern. Also, the Fleiss Kappa rate has not been modified to get better performance in the hate detection area in Twitter. We applied the most famous Natural Language Processing algorithms on both our and Ziems annotated datasets and compared the results. After tuning the Fleiss Kappa number and balancing the dataset, we achieved our best result: an F1-score of 0.85 by the Bert model.

We aim to compare fully supervised, fully unsupervised, and semi-supervised methods on our dataset. We also plan to annotate the same data with annotators of other ethnicities to see how the results differ. In addition, we will investigate whether the hybrid algorithm with XGBoost and Grey wolf optimization will improve the results.

**Committee:**

- Professor Ronak Etemadpour, Mentor, The City College
- Professor Sarah Ita Levitan, Hunter College
- Professor Zheng Peng, The City College