AIMS^2 Research Project in Computer Science

Research Duration:　　　　Summer 2020 (June – August 2020)

Faculty:　　　　Kyle Dewey

Email address:　　　　kyle.dewey@csun.edu

Contact No:　　　　JD 4419, x4316, Office hours via Zoom (email me times
　　　　and personalized link)

Title of Project:　　　　Sentiment Analysis of Tweets

Project Background: In online social networking, there is often a need to perform moderation, be it by blocking particular posts, or potentially whole users. Due to the sheer quantity of posts and users, performing this moderation fully manually is often not feasible. To this end, we need automated tools which can assist human moderators. Towards assisting with this moderation, *sentiment analysis* is of great value. Sentiment analysis is used to automatically determine if a particular post displays a positive, negative, or neutral attitude.

Goals and Objectives of the Project: The goal of this project is to build a sentiment analysis system for Twitter tweets. This system will utilize machine learning for this purpose. This system will be built in a stepwise fashion:

1. An appropriate *training set* will be identified, containing posts which have already been labelled with their given sentiment. This may involve using Twitter's API to load data.
2. A machine learning model will be trained on this training set, and will attempt to learn the patterns of positive/negative/neutral sentiment tweets.
3. We will then evaluate the model's performance, and potentially revisit the prior step as needed to make the model more robust.
4. Potentially, to enable moderators to easily use the model, we may embed it in a browser plugin which will dynamically display the sentiment of tweets viewed online.

Project Outcomes: Students will gain hands-on experience working with machine learning and using related tools (e.g., scikit-learn, Keras, etc.). Depending on the project direction, students may also gain familiarity using web APIs, and writing browser plugins.

Expectations from Students: Students will meet virtually via Zoom ~1 hour per week to learn necessary machine learning background and to provide project status. All communication will be via free software, namely Zoom (video), Slack (text), and GitHub (project management). Additional reading may be needed. Students will use Python, which is commonly used in machine learning. Incoming students must have taken at least a CS1 equivalent (e.g., CSUN's COMP 110/L). Prior Python experience is ideal, but not required. No prior experience with machine learning is necessary. Students must provide their own computer; no special software or hardware is needed. **After the project**, students are expected to make an oral presentation and present a poster at the AIMS^2 research symposium, tentatively in Spring 2021.