**Analyzing the Sentiment of Twitter Users Towards the Israel-Palestine Conflict**

Group 7: Quan Ly, Michelle Odnert, Kaiden Ong, Jonathan Osorio, Stephanie Wang

**Motivation:** Our group wanted the chance to take the machine learning concepts we’ve covered in class and apply them to something relevant to our lives. We decided to conduct an analysis of the sentiment of the public towards the current state of the Israel-Palestine conflict as this is a topic that has been extremely relevant in recent weeks and is important to expand our knowledge on. This question comes from the social science application side of data science and will involve the machine learning concepts of sentiment analysis and natural language processing.

**Question:** With our analysis, we want to ask what is the general consensus on the current situation with Israel and Palestine? Are more people pro-Palestine? Pro-Israel? Or are they neutral? As well, has sentiment changed over time? To better understand this, we want to analyze Twitter analytics to understand the general public’s point of view especially since there are similar studies that have already been done.

**Data:** We will acquire our data using an API called Octoparse to scrape Twitter. We scraped Twitter under the hashtag #IsraelPalestineConflict (and possibly more to be added), looking at tweets from the date range of October 7, 2023 (the date of the attack from Hamas) to the present day. The data was exported to a CSV file where we currently have 273 rows. We have columns on the date posted, who posted it, what the tweet contains, how many likes it received, how many reshares it received, and the number of looks it received. We will scrape more in the coming days potentially with different keywords or more demographic information. From these top 273 tweets in the last month, they have seen a total sum of 5,440,291 views and 37,991 retweets. This gives an immediate insight into how broad an audience just a few tweets can reach. [Link to our data](https://drive.google.com/file/d/1Fu22KSrVNu3_Bl7JHq_aYNR77t1uCfK_/view?usp=sharing)

**Related Work:** Other projects that have been completed around our project include general sentiment analysis studies from other social media platforms including Twitter and other topics like ChatGPT. In the paper ‘Sentiment Analysis on Twitter Data’, a general sentiment analysis study on Twitter data using machine learning algorithms was conducted. The study categorizes tweets into positive, negative, or neutral sentiments related to specific query terms, enabling applications to assess customer feedback for product improvement (Sahayak et al., 2015). Additionally, there has been a project done more specifically on analyzing Twitter tweets from December 2022 to January 2023 about ChatGPT (Huang, 2023). Three research questions were asked regarding the main topics and sentiments of the conversations around early ChatGPT users. The study was used to allow us to understand and assess ChatGPT's capability, effectiveness, and challenges. A third study was done similarly using the same sentiment analysis and machine learning approaches but delved into users’ perceptions and opinions regarding electronic products like mobiles and laptops (Rajasree & Neethu, 2013).

**Approach and Analysis:** We will use the scraped tweets to train and test several machine-learning models. Our goal with the models is to understand and contextualize the current Twitter sentiment on the Israel and Palestine conflict. We plan to use models such as decision trees, kNN, and others learned in class to categorize tweets into groups such as pro-Israel, pro-Palenstine, or neutral. The machine learning models’ predictions will create a confusion matrix. From that confusion matrix, we plan to analyze and measure the performance of our models using statistics such as accuracy, precision, recall, etc. The first analysis that we plan on performing is identifying which models perform best for sentiment analysis and tweet classification. Using the results from the first analysis, we can explore how sentiment has changed over these past few months as the war has started and raged on. Lastly, we hope to get location/geographical data from Twitter scraping. With this information, we can create heat maps and see how people around the world feel about this ongoing war. Other plausible analyses include, Twitter following influence (how users with more followers might shape public opinion) and shifts in sentiment in correlation with events/attacks.

**Plan:** To ensure the successful completion of this project by the end of the quarter, our goal is to have regular in-person meetings at least once a week to discuss our progress and share our discoveries. Task assignments will be flexible, adapting to individual schedules, skills, and comprehension of the subject matter. At the present moment, Michelle will be responsible for extracting data from Twitter and creating a dataset for the team's use. Subsequently, everyone will share the responsibility of performing fundamental data analyses. Each team member will be assigned specific machine learning models to explore, work with, and experiment with. Kaiden will assume a leading role in addressing any coding challenges that may arise. Assuming everything proceeds as planned, each team member should be able to evaluate their work's outcomes and provide valuable insights to the rest of the group. Stephanie will take the lead in composing the report summarizing our findings. Lastly, Johnny and Quan will be in charge of preparing the presentation slides, with contributions from every team member during the presentation. Given the significant population size and the country's international connections, we anticipate the results to exhibit a pro-Israel bias.

**References**

Huang, H. (2023, May). Twitter Sentiment Analysis about ChatGPT. GitHub. https://github.com/hxycorn/Twitter-Sentiment-Analysis-about-ChatGPT

Rajasree, R., & Neethu, M. S. (2013). Sentiment Analysis in Twitter using Machine Learning techniques. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*.

Sahayak, V., Shete, V., & Pathan, A. (2015, January). Sentiment Analysis on Twitter Data. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, *2*(1).