**Project Plan**

**Team Members:**

María Tapia Costa

Mohamed Arafaath

**Project Title:**

An In-depth Analysis of Public Sentiment: Reddit Discourse on the Israel-Palestine Conflict

**Abstract:**

The project focuses on conducting a comprehensive analysis of public sentiment expressed within Reddit discussions pertaining to Palestine-Israel conflict. Using various NLP methodologies, the project aims to elucidate the dynamic evolution of opinions, discern influential variables, and scrutainize potential biases prevalent in the discourse.

**Envisioned data to be used:**

[Kaggle](https://www.kaggle.com/datasets/asaniczka/reddit-on-israel-palestine-daily-updated)

**Envisioned algorithm(s)/method(s) to be used:**

1. **Named Entity Recognition (NER):** Identify key services/people/organisations mentioned in conversations, and track their breadth and context.
2. **Topic Modelling:** Using LDA to uncover prevalent topics and their evolution over time.
3. **Analysis of Linguistic Features:** Using POS tagging to study grammatical trends, potentially correlating them with sentiment categories.
4. **Trend Analysis:** Using TF-IDF or word embeddings to identify and track key terms and phrases associated with different sentiment categories understanding their significance.
5. **Sentiment Analysis:** Measure the emotional tone of conversations over time with all the analysis carried out previously.

**Overview of Evaluation and comparision strategy:**

We can use perplexity, coherence, v-score to evaluate the LDA model to assess the interpretability of the topics, ensuring that they effectively capture distinct sentiments

1. **Sentiment Distribution:** Analyze the distribution of sentiment (Pro Israel, Strongly Pro Israel, Pro Palestine, Strongly Pro Palestine and Neutral) within each topic to understand the predominant sentiment within each category.
2. **Sentiment Change Over Time:** Track how the sentiment for each topic changes over time to identify shifts in public opinion or sentiment trends.
3. **Keyword Analysis:** Conduct keyword analysis within each topic to identify key terms or phrases associated with different sentiments, helping to differentiate between the categories.