## **CSCE 5290 Natural Language Processing**

**Project Proposal: Twitter Emotion Classification**

Dataset: [Emotion Dataset For Emotion Recognition Tasks](https://www.kaggle.com/datasets/parulpandey/emotion-dataset)

**Title: Twitter Emotion Classification**

**Members:**

Anusha Chilakamarri -11664602

Ram Srinivas Katragadda – 11664589

Saiteja Goud Akkala – 11703752

Jeevan Parakala – 11706336

Githublink- <https://github.com/ram2312/NLPProject.git>

1. **Motivation:** The development of Transformers has lately eclipsed the field of natural language processing. Transformers have no preference whatsoever for traditional sequence-based networks. For sequence-based tasks like text creation and text categorization, RNNs are the first line of defense. However, the problem of encoding long-term dependence in the text was overcome with the introduction of LSTM and GRU cells. However, since the model cannot learn in parallel, learning it using LSTM cells is a challenging process. Their findings may help corporations and governments implement programs, events, and goods. They may also be useful for marketing, gathering consumer feedback, and monitoring public opinion.
2. **Significance:** Social media traction is crucial for companies. Strong social media strategies help brands stand out from the competition online, especially with younger consumers. Because of its features for trending topics and mentions, Twitter is particularly helpful. Because it enables you to see how consumers truly feel about your brand, this data is helpful. It is simpler to comprehend your brand's current visibility and your desired changes when you have access to candid input. A business or individual will be able to adapt to fit its brand based on facts with this understanding following investigation.
3. **Objectives:**

* Create a multiclass classification model to classify tweets into the following six categories: surprise, fear, rage, joy, sadness, and love.
* Track employee satisfaction levels, find out what customers are saying about the competition, and forecast market trends for specific company niches.
* Analyze the model's performance using relevant measures, such F1-score, accuracy, precision, and recall.
* Give insights on the emotions that are common in Twitter chats and how they affect various sectors.

1. **Features:**

* The application of cutting-edge deep learning models for emotion categorization, including CNN, Transformer, and LSTM.
* The creation of an interactive web interface enabling users to examine the findings of emotion analysis on Twitter data.
* Pre-trained word embeddings are integrated to improve classification performance by capturing semantic information.
* Using data augmentation strategies to improve the resilience and generalization of the model.

1. **Dataset:**

* We will utilize the "Emotion Dataset for Emotion Recognition Tasks" available on Kaggle (<https://www.kaggle.com/datasets/parulpandey/emotion-dataset>).
* The collection of tweets in the dataset has been annotated with six distinct emotions: anger, fear, joy, love, sadness, and surprise.
* Tokenization, URL removal, and special character removal are among the preprocessing procedures that will be carried out.
* The sizes of the datasets are as follows:
  + Validation data: 2000 samples
  + Train data: 16000 samples
  + Test data: 2000 samples.

1. **Visualization:**

* To assess the model's performance and show the distribution of emotions in the dataset, we will use visualizations like bar charts and confusion matrices.
* Furthermore, we want to create interactive visualizations inside the web-based interface, which will enable users to dynamically examine the findings of the emotion analysis.
* Below is the workflow.

