**Project Proposal:** Real-Time Emotion Recognition from Speech

**Team Members:** [Jayesh Locharla – JL23BM] [Sruthijha Pagolu - SP23BU]

**Course:** CAP 5619 - Deep and Reinforcement Learning, Spring 2024

**Department:** Computer Science, Florida State University

**Introduction:**

In the digital era, where human-computer interaction is becoming increasingly prevalent, understanding human emotions through speech can significantly enhance the quality of interaction and user experience. Our project aims to implement a real-time emotion recognition system using deep learning techniques to classify emotional states from speech. This system has vast applications, ranging from improving customer service experiences to supporting mental health assessments and creating more interactive AI systems.

**Objectives:**

1. **Primary Objective:** To develop a real-time system capable of recognizing and classifying various emotional states from human speech.
2. **Accuracy Goal:** Achieve a classification accuracy of at least 95% on the RAVDESS dataset.
3. **Real-Time Processing:** Ensure the system can process and classify speech in real-time, with minimal latency.
4. **Robustness:** Make the system robust against background noise and variations in speech, such as accents and intonations.

**Proposed Methodology:**

1. **Dataset:** Utilize the RAVDESS dataset, which contains audio recordings labeled with various emotional states. The dataset will be augmented with noise, pitch variations, and speed changes to improve the robustness of our model.
2. **Feature Extraction:** Implement feature extraction techniques, focusing on Mel-Frequency Cepstral Coefficients (MFCCs), Zero Crossing Rate (ZCR), and Root Mean Square Energy (RMSE), to capture the nuances of emotional expression in speech.
3. **Model Architecture:** Use Long Short-Term Memory (LSTM) networks to model the temporal dynamics of speech features effectively. The LSTM model will be trained to classify different emotions based on extracted features.
4. **Real-Time Implementation:** Develop a pipeline for real-time audio processing and classification. This involves capturing live audio input, extracting features in real-time, and using the trained LSTM model to predict the emotional state.

**Expected Outcomes:**

Upon completion, we expect to have a fully functional real-time emotion recognition system that can accurately classify emotional states from speech. The system will be evaluated based on its accuracy, latency, and robustness under various conditions.

**Conclusion:**

This project not only aims to advance the field of speech emotion recognition but also offers practical applications in enhancing human-computer interactions across various domains. By achieving our set goals, we hope to contribute to the development of more empathetic and responsive AI systems.