**Group Members:** Eric Chao, Krish Sapru, Kevin Guo

**Problem Statement:** Diving deeper into classifying audio files using Mel Spectrograms and Convolutional Neural Network based on sentiment and dialogue.

**Project Final Report:**

For our final project, we have created an audio sentiment analyzer that uses machine learning to determine the sentiment of an audio clip. The inspiration for this project came from the idea of other countries incorporating an audio sentiment analyzer to rate social credit scores. There can also be many applications of this project, including using it in voice assistants such as Siri, using it to detect signs of depression, and using it for therapy.

Our process of doing the project consisted of three steps. First we gathered audio data that were categorized into positive and negative sentiments and created directories for their respective categories. Then, with help from professor Snyder, we cut all the audio files to be the same length and created mel-spectrograms out of the audio clips. Finally, we fed the mel spectrograms from each directory into a CNN (Convolutional Neural Network) in order to train it to properly analyze subsequent mel spectrograms.

The basic idea of the Neural Network is that it will take all the mel spectrograms from a directory and create an image that best represents the positive and negative sentiment, so that when we feed it another audio clip, it can tell us if it is more alike the mel spectrogram of the positive or negative sentiment. The technologies we used for this project include Tensorflow, Librosa, Numpy, Pandas and Keras.

Our final result was a Machine Learning algorithm that outputted decimal values that determined how positive or negative an audio clip was, if the values were closer to 0, they were considered negative and if they were closer to 1, they were positive. The models we trained displayed different results through collaboration with the neural network. For the data presented in audio signals, we separated and coagulated different sections to display positivity or negativity in an array. For example, a signal could show [positive, negative, positive]. If we wanted to get the average sentiment of the full clip, we would find the average of all the values in the array.

Throughout the project, we overcame many hurdles. Our first big hurdle was the proper collection of data. At the time, we were debating between using two big audio datasets, one which came from the MELD dataset and had sentiment tags for each audio clip which we indexed based on our needs into four categories: DIALOGUE\_ID, UTTERANCE\_ID, Dialogue, and Sentiment, and one that came from a study conducted by researchers at Carnegie Mellon that had audio clips of various people doing movie reviews. We later learned that we couldn’t use the dataset with the movie reviews to feed the CNN since they did not have positive/negative sentiment tags. Both datasets were too large for github to process so we shared the datasets through google drive and picked out a smaller set of around 1000 audio clips to feed the CNN. Our next hurdle was creating the Mel spectrograms, which was built with the help of Professor Snyder’s notebook and digesting the reading “Hands on Machine Learning with SciKit and Keras.” Subsequently, we used a CSV file to display and sort our findings into arrays that we indexed based on what the neural network required to train. We compared our trained data to our tested data which was later used to display the findings in the Jupyter Notebook.

In the future, we plan to continue looking for more datasets that have positive/negative sentiment labels so we could feed the network to get more accurate results. We would also experiment with feeding different types of data into the CNN, for example, we would feed directories that would be categorized based on emotion so the CNN could analyze patterns for surprise, fear, anger, sadness and joy.

**Datasets & Digested Readings:**

* Emotion recognition dataset: <https://github.com/declare-lab/MELD>
* [**https://www.knowledgeisle.com/wp-content/uploads/2019/12/2-Aur%C3%A9lien-G%C3%A9ron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow\_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O%E2%80%99Reilly-Media-2019.pdf**](https://www.knowledgeisle.com/wp-content/uploads/2019/12/2-Aur%C3%A9lien-G%C3%A9ron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O%E2%80%99Reilly-Media-2019.pdf)
* Data + ipynb file with results:
  + <https://drive.google.com/file/d/1gTFpvTVBILXOcZSsq8ooTrgCvcDekNs0/view?usp=sharing>