Speech Sentiment Analyzer

Marcelino Bonamutial 2501965140, Alvian Daniel Susanto 2501968924, Daniel Aditya Tumansery 2501985350, Emmanuel Brandon Hamdi 2501970941, Nathaniel Susianto Sutanto 2540127616.

# Overview

The project that we are planning to do is Sentiment Analysis. Speech Sentiment Analysis is a study of a person's emotion or behavior towards a conversation, or in general such as what they think of something [1]. The scope of our project will involve using a dataset that is already provided by Kaggle through this link [https://www.kaggle.com/datasets/imsparsh/audio-speech-sentiment.](https://www.kaggle.com/datasets/imsparsh/audio-speech-sentiment.%20) and also <https://www.kaggle.com/datasets/uldisvalainis/audio-emotions> Our expectation of the product that we are making is that the input will be a .wav file, either from the dataset such as the test set, or manual recording and converted to a .wav file. The output will be as follow “Sentiment: {Negative/Neutral/Positive}”.

# How the Product Works

Our program will accept audio or human voice as input. Then, the program provides output in the form of text which is the result of sentiment analysis and classification. The possible output classes are negative, neutral, and positive.

# Methodology

This section describes the technical detail of the product. Put the workflow of your product or research in the form of a chart

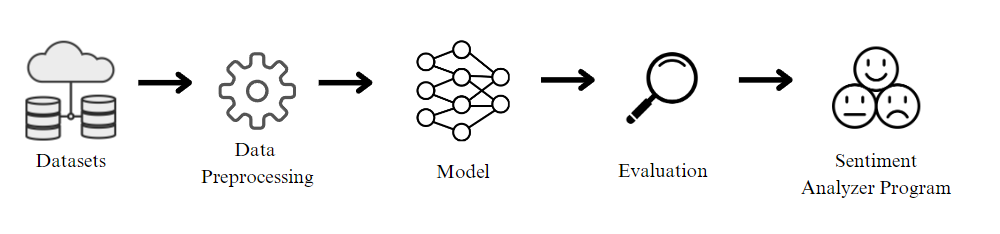


Figure 1 Program Building Workflow Chart

## Model

The Image above is the methodology workflow. Starting from getting the dataset from Kaggle. Then creating MFCC (Mel Frequency Cepstral Coefficients), where we use the function to extract the spectral characteristic of the audio signal, specifically the shape of the spectrum. Then we preprocess the audio using MFCC that we already made by normalizing the audio and trimming or padding the audio to make the audio the same size. Then we split the data with a train and test ratio of 80:20.

After that, we preprocess the data again using a padding sequence and one hot encoder. The one hot encoder is used for converting data type into a numerical one, Then we created the model architecture using Convolutional Neural Network (CNN), a widely used deep learning model that is used in many images or speech classification tasks [2].

We create the model architecture in sequential order, by first adding the dense layer consisting of 256 units using relu as the activation function as the input layer. LeakyRelu or Leaky Rectified Linear Unit is commonly used in a neural network for audio, where positive values are passed in and the negative values are set to zero. Then a dropout layer with a dropout rate of 0.2 after the first layer to prevent overfitting. The second layer used 128 units using LeakyRelu as the activation function. Then another dropout rate of 0.2 is added after the second layer, and after that, a third layer with 64 units and Relu activation function is added with another 0.2 dropout layer, a dense layer with 32 units with Relu, and a 0.2 dropout layer is added also.

Finally using units equal to the number of multiclass in the problem using softmax as the probability distribution over the classes, where each value is the likelihood to the belonging of a specific class. The model is compiled using the categorical cross-entropy loss function, the optimizer is using Adam, and the metric uses accuracy to predict the correct samples. The model was then fitted with a batch size of 32 and an epoch of 10. The model then is evaluated using loss and accuracy for the test data.

The implementation then uses the input of a recorded audio file, then the audio is preprocessed and then predicted using the created model, using the sentiment labels such as Positive, Neutral, and Negative to predict it.

## Dataset

The first dataset used is titled ‘Audio Speech Sentiment’ and is sourced from Kaggle [3]. The dataset contains speech audio files, their corresponding spectrograms, and sentiment labels in a CSV file. Audio from this dataset is all speech in English that has been classified into 3 classes: neutral, positive, and negative. All speech audio also seemed to be spoken with a South Asian/Indian accent which could make it more difficult to use on other speech files of other dialects. In total, there are 250 speech audio in the training file and 110 in the test file, unfortunately, the test audios are not labeled, so only the train audio will be used on this project. The audio labels are also equally distributed on the train data with 35% being negative, 33% positive, and 32% neutral.

The second dataset that is used is titled ‘Audio emotions’ and is also sourced from Kaggle [4], the dataset contains 35685 audio that has been classified into 7 different classes which are angry, disgusted, fearful, happy, neutral, sad, surprised, from the 7 classes of the datasets we turned it into the 3 speech sentiment with the happy class turned into positive, then the angry, disgusted, fearful and sad classes are turned into negative, after that, the neutral and surprised classes are turned into neutral.

# Result

This section displays the result of your project.

* The result of the model training is as on Figure 2 below.

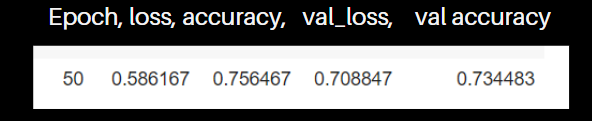


Figure 2 CNN Model Result

* The sentiment analyzer program is as shown as on figure 3 where first the audio is recorded then saved, after that the sentiment is predicted using the CNN model.

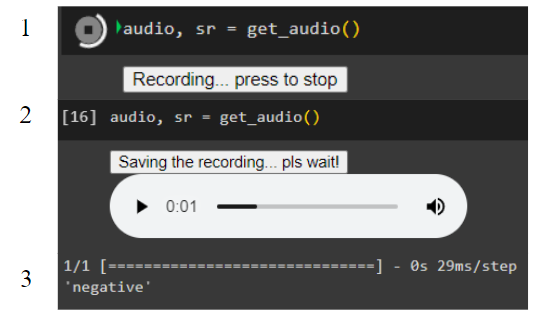


Figure 3 Speech Sentiment Program

# References

|  |  |
| --- | --- |
| [1] | S. Maghilnan and M. R. Kumar, “Sentiment analysis on speaker specific speech data,” in *2017 International Conference on Intelligent Computing and Control (I2C2)*, pp. 1–5. 2017. |
| [2] | U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu, and G. Karthick, “Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM),” *Wireless Pers Commun*, 2021. |
| [3] | S. Gupta, “Audio Speech Sentiment,” 2021. [Online]. Available: <https://www.kaggle.com/datasets/imsparsh/audio-speech-sentiment> [Accessed May. 28, 2023]. |
| [4] | U. Valainis, “Audio emotions,” 2020. [Online]. Available: <https://www.kaggle.com/datasets/uldisvalainis/audio-emotions> [Accessed May. 28, 2023] |