Audio Speech Commands

Kathy Simon | Data Scientist

Table of Contents

Problem Statement

Approximately 15% of adults report hearing loss. Video conferences, TV, streamed shows and movie theaters provide closed captions. But conversations and live events can be difficult to hear. My ultimate goal is to create a model that can provide a transcript of a live conversation. Eventually implementing an app that can provide a transcript that can be followed at a live event .

As proof of concept, I will create model that takes audio sound and predicts speech commands.

- Dataset: TensorFlow: Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition By Pete Warden
- Classification Model of Convolutional Neural Networks
- Metrics of success based on accuracy

Background Information on Hearing Loss

Affects of Hearing Loss

- Mental Health
- Cognitive Decline
- Relationship Issues

Types of Hearing Loss

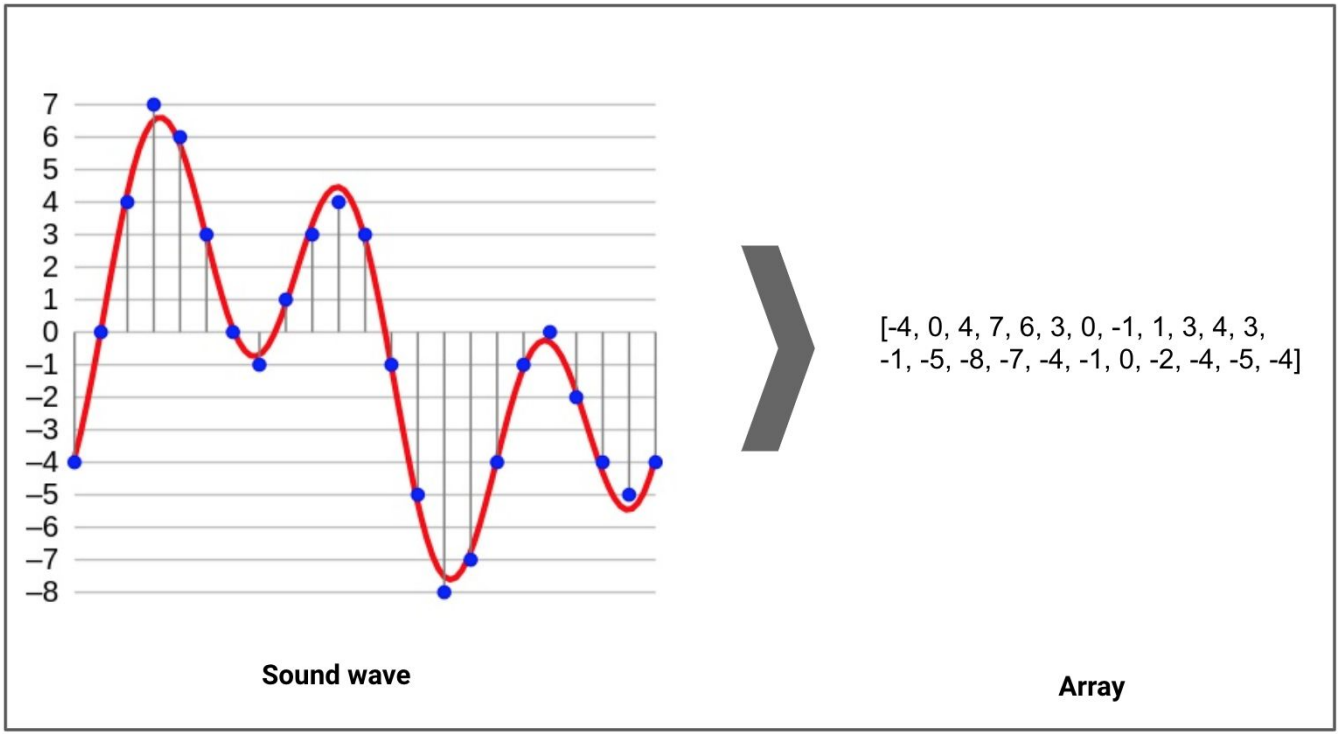
- Conductive
- Sensorineural
- Mixed Hearing Loss

Limitation of Hearing Aids

- Do not restore normal hearing
- Amplify all sounds!
- Adjustment period
- Cost Prohibitive

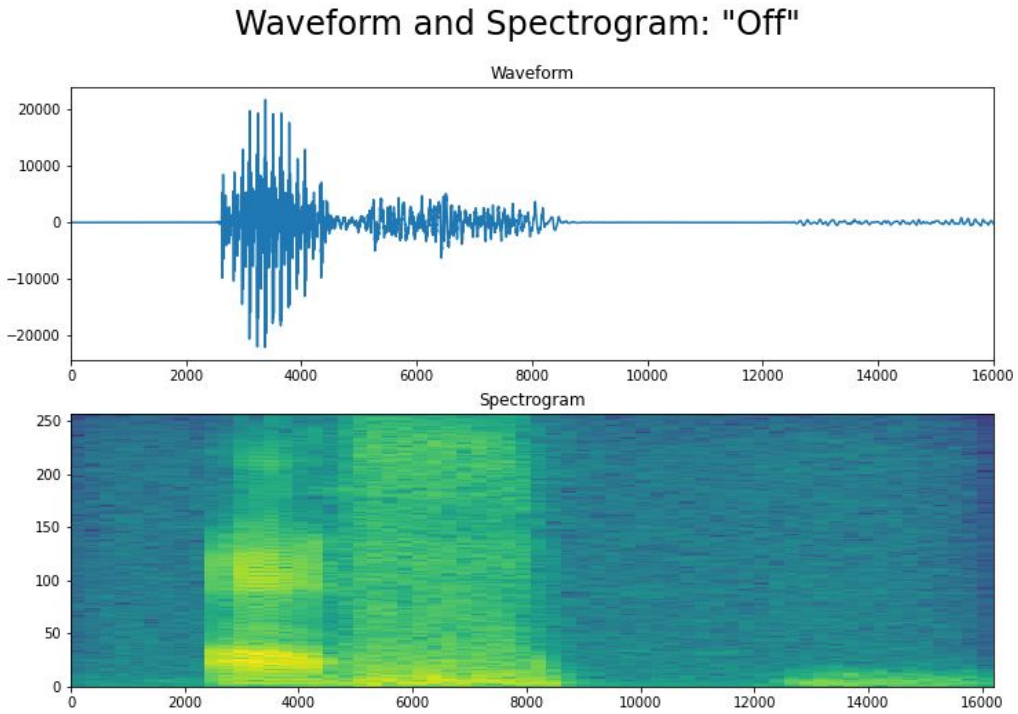
https://www.carsonhearing.com/hearing-loss-articles/these-4-side-effects-of-hearing-loss-will-surprise-you/
https://www.hopkinsmedicine.org/health/conditions-and-diseases/hearing-loss/types-of-hearing-loss
https://www.fda.gov/medical-devices/hearing-aids/benefits-and-safety-issues

Background Information on Audio



http://www.digitizationguidelines.gov/term.php?term=waveformsound
https://pudding.cool/2018/02/waveforms/

Waveform and Spectrogram

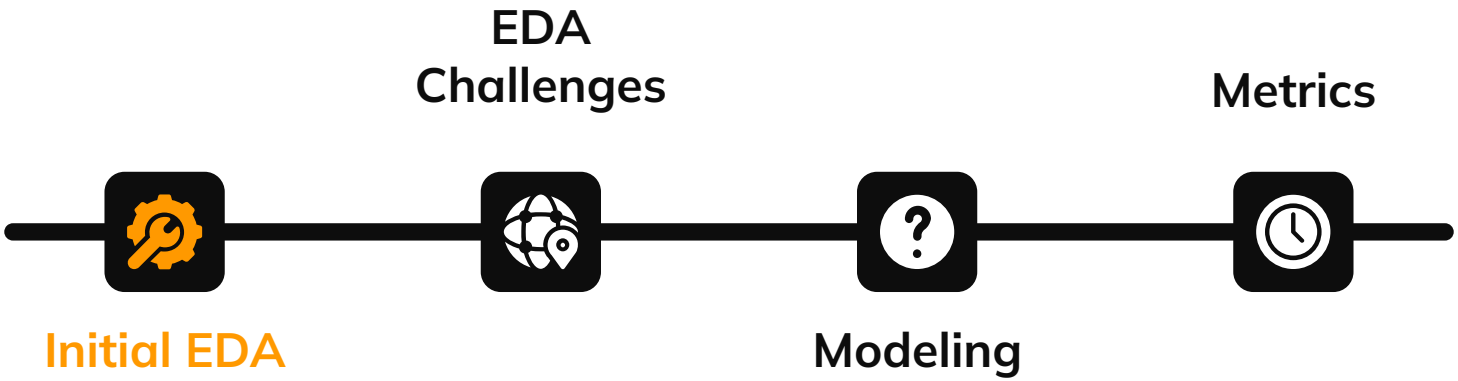


A **waveform** describes the pattern of sound pressure (amplitude) overtime.

A **spectrogram** is a visual of how the frequencies of the audio vary overtime.

http://www.digitizationguidelines.gov/term.php?term=waveformsound
https://en.wikipedia.org/wiki/Spectrogram

Audio Speech Commands – Workflow



Speech Command Data

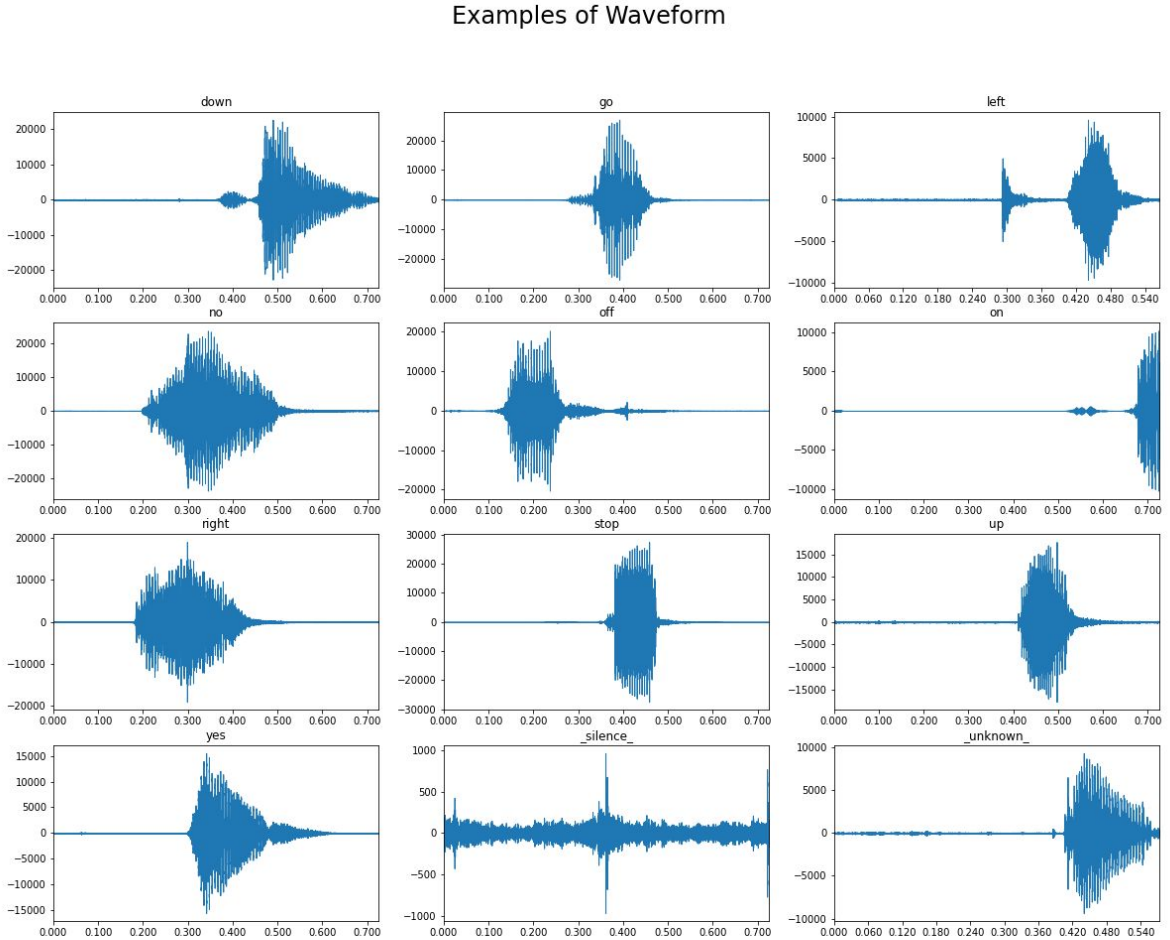
- Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition By Pete Warden
- 1 second .wav audio files
- Amount of data
  - train: 85,511
  - validation: 10,102
  - new data: 4,890

Speech Commands

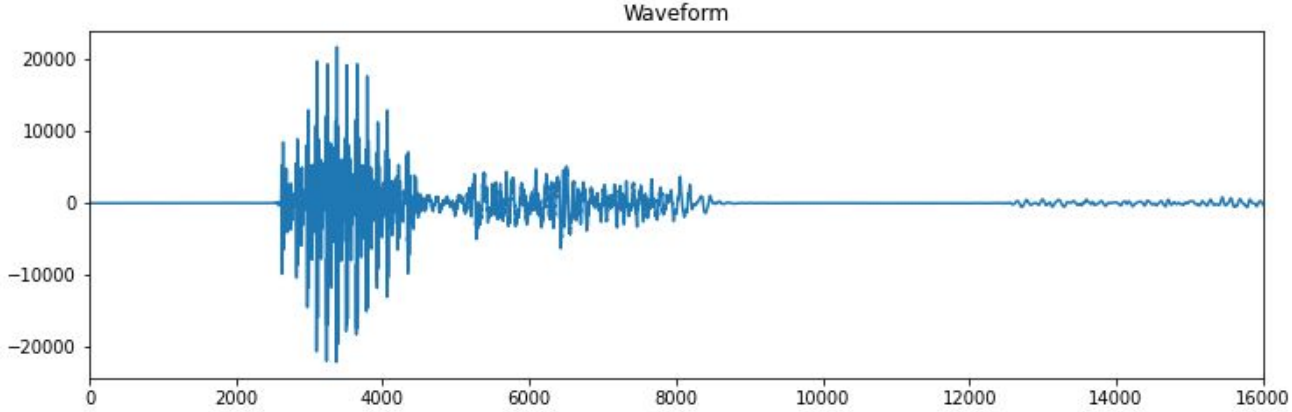
- Down
- Go
- Left
- No
- Off
- On
- Right
- Stop
- Up
- Yes
- _silence_
- _unknown_

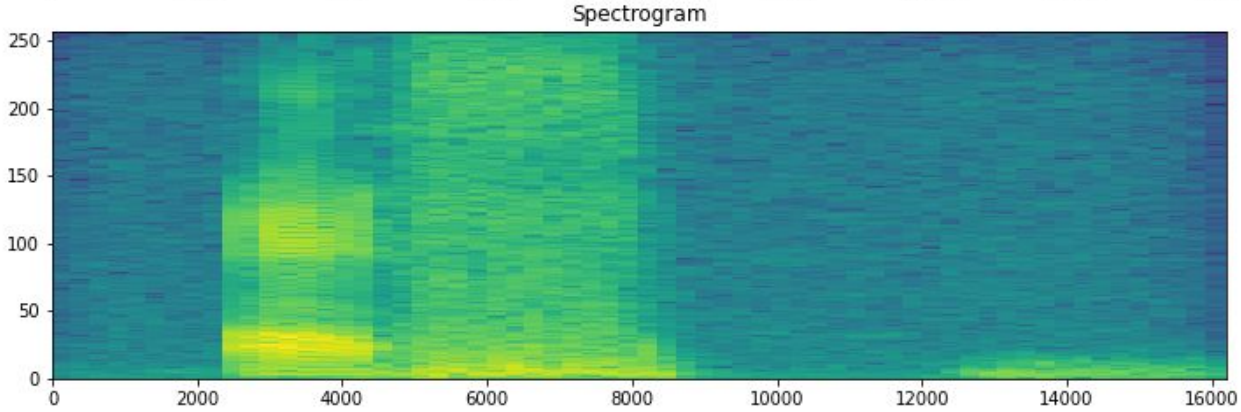
Example of Data

- Down
- Go
- Left
- No
- Off
- On
- Right
- Stop
- Up
- Yes
- _silence_
- _unknown_

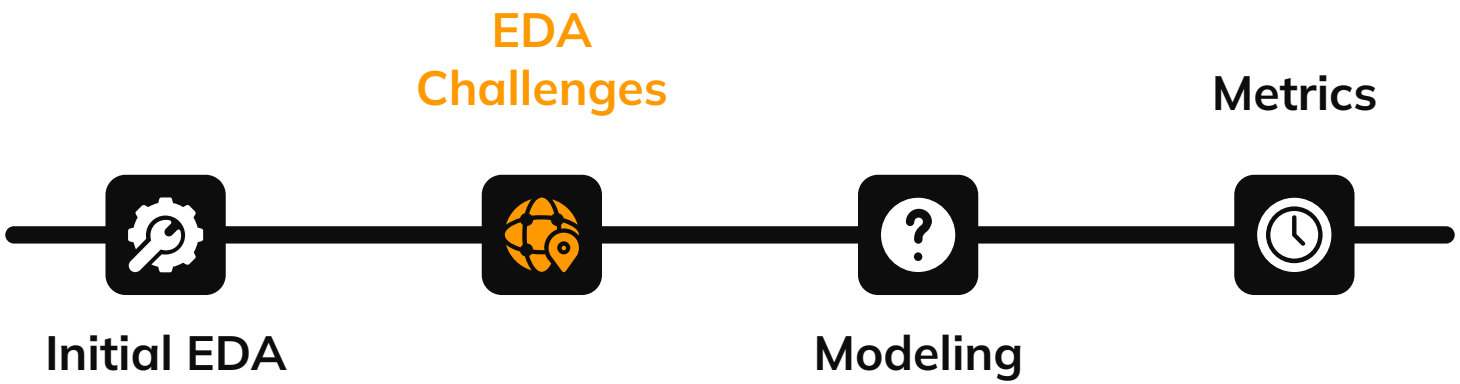


Example of Data – Waveform and Spectrogram: "Off"





Audio Speech Commands – Workflow



EDA Challenges

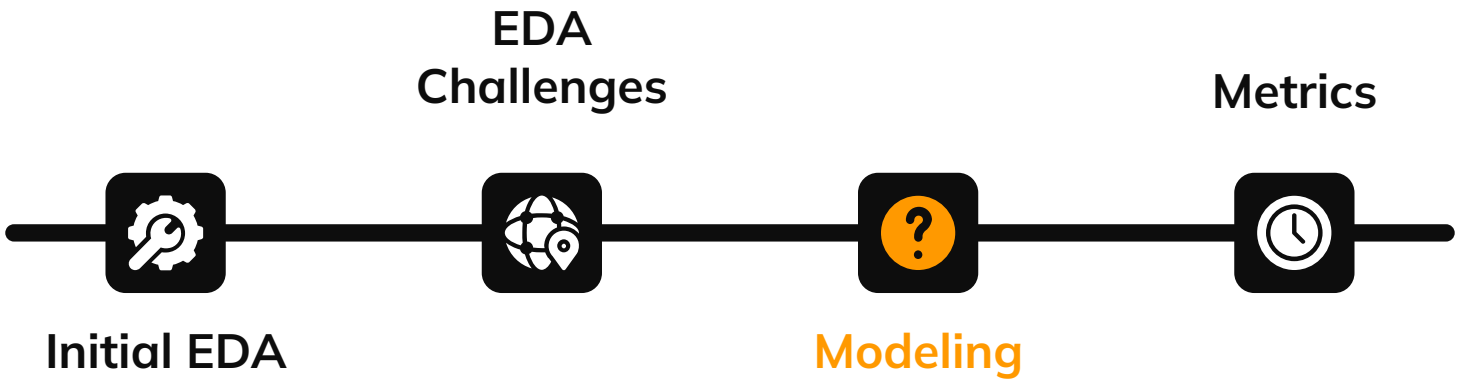
Limitations

- Size of Data
- Tensorflow
- Google Colab

Data

- Spectrogram images have to be all the same sizes.
- Images need to be padded to ensure this.

Audio Speech Commands – Workflow



Information about Convolutional Neural Networks (CNN)

1. Neural Networks (NN) is a collection of layered algorithms that takes an input, performs different transformations and produces an result.

2. CNN are more efficient than other NN

3. CNN is well suited to deal image data, including the audio spectrograms

Increasing Performance when Using CNN in TensorFlow

1. **Prefetching**

   tf.data.AUTOTUNE

2. **Parallelizing Data Extraction**

   tf.data.AUTOTUNE

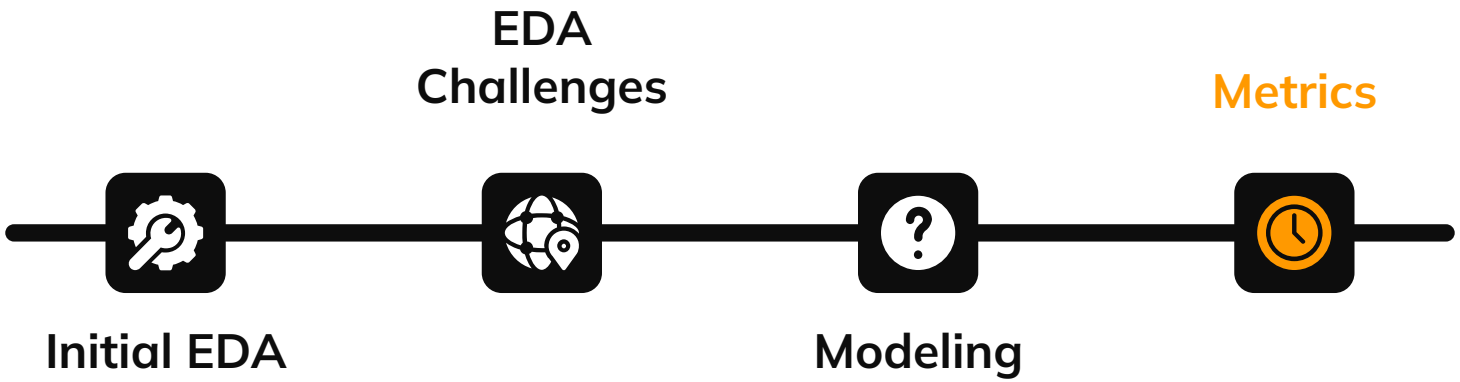
3. **Parallelizing Data Transformation**

   tf.data.AUTOTUNE

4. **Caching**

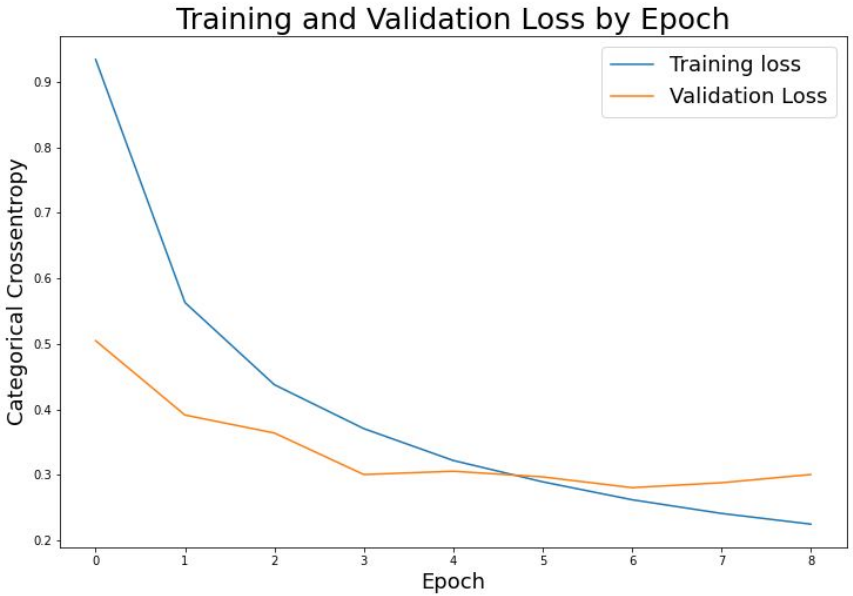
   tf.data.Dataset.cache

Audio Speech Commands – Workflow

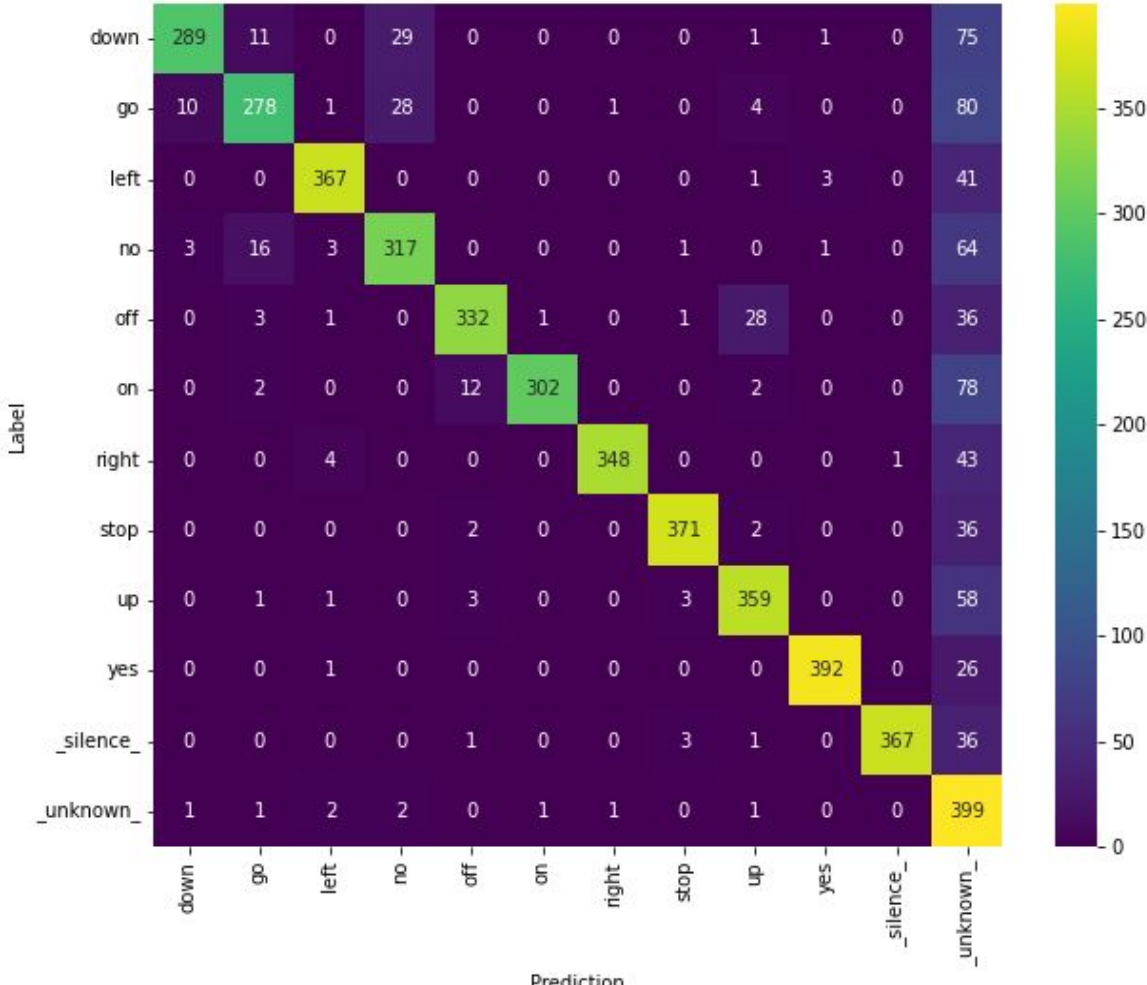


Model Summary on Accuracy

	Accuracy
Baseline	63%
CNN Training	93%
CNN Validation	93%
CNN New Data	83%



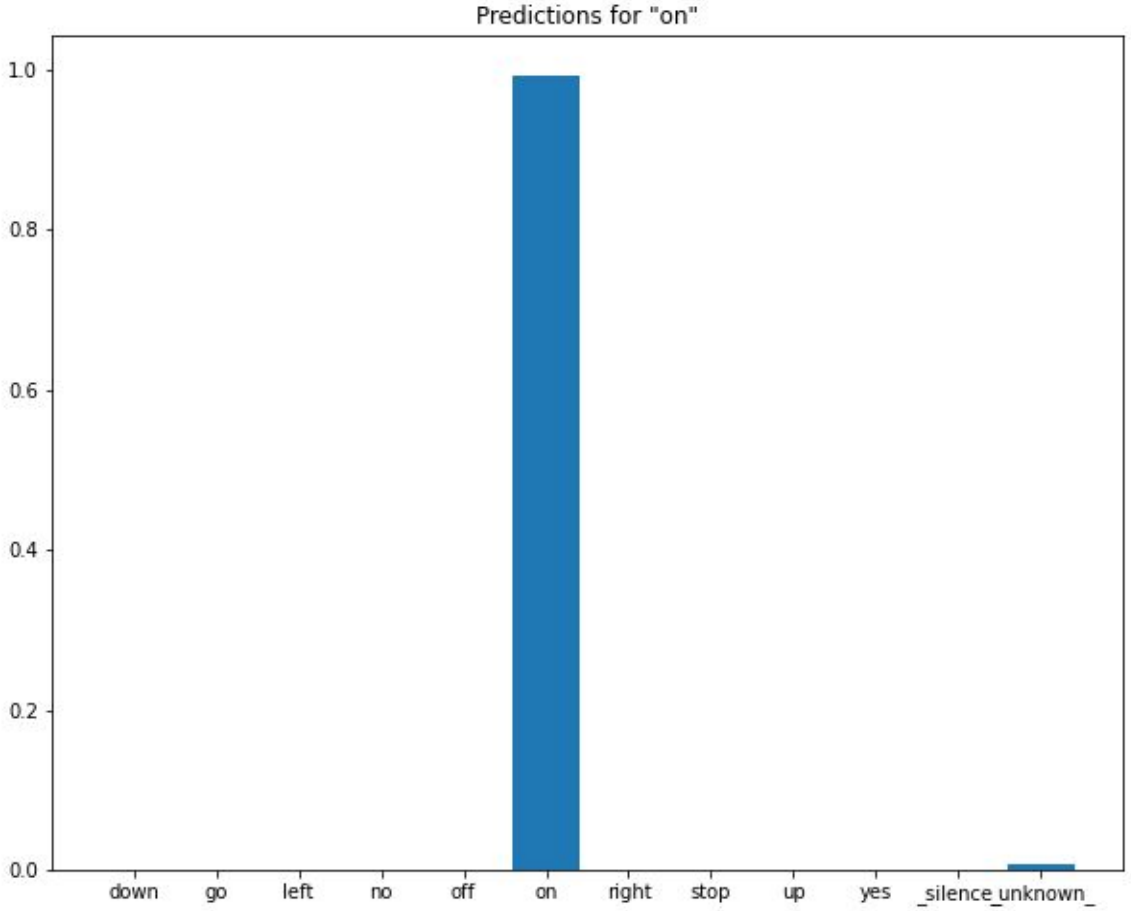
Confusion Matrix



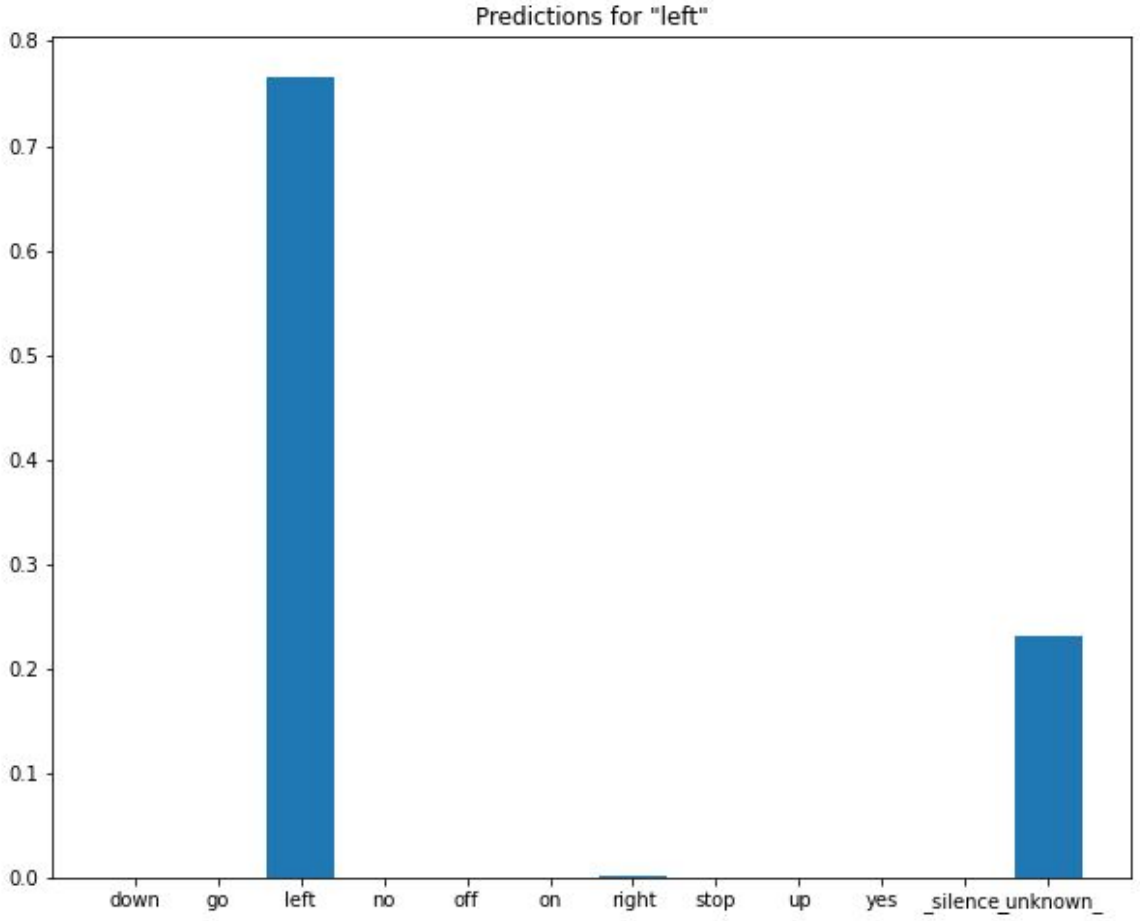
Most Common Misclassification

- Down : _unknown_, no
- Go : _unknown_, no
- No : _unknown_, go
- Off : _unknown_, up
- Left: _unknown_
- On: _unknown_
- Right: _unknown_
- Stop: _unknown_
- Up: _unknown_
- Yes: _unknown_
- _silence_ : _unknown_

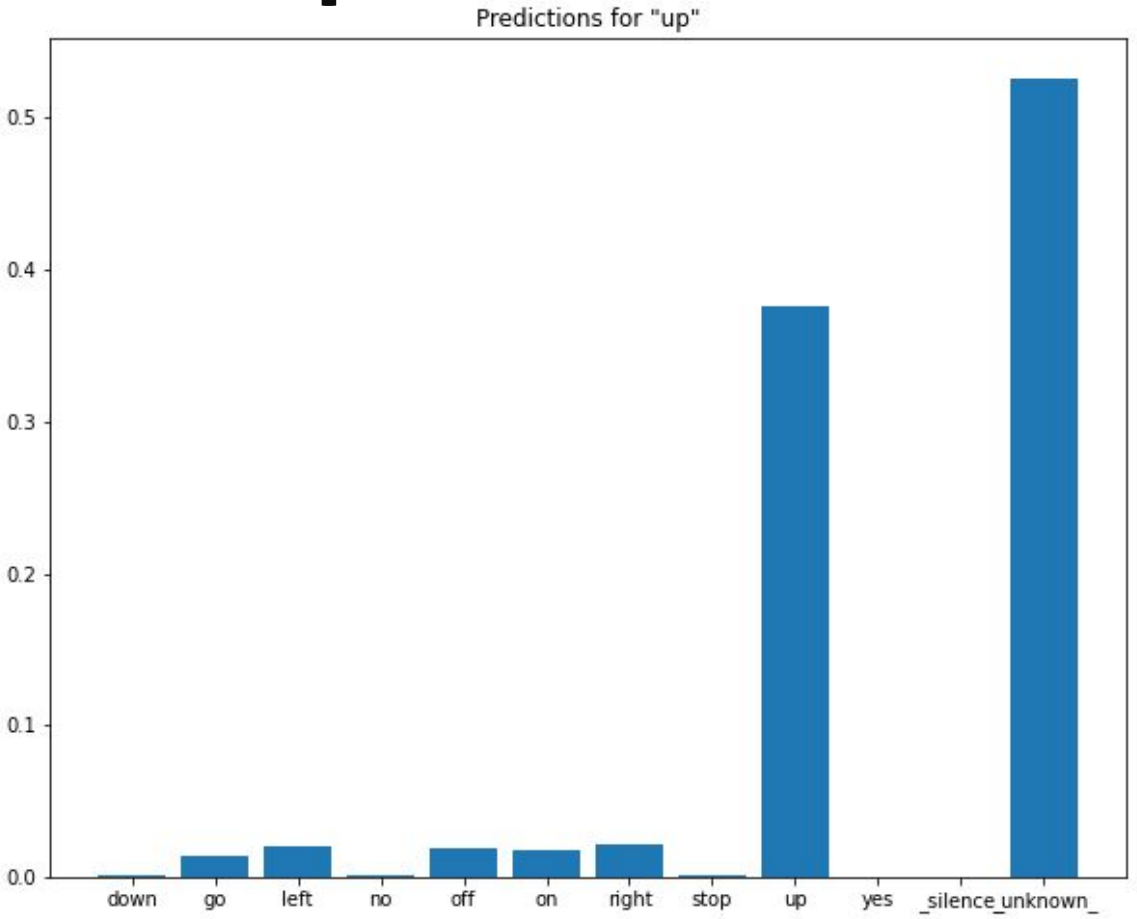
Predictions for “on”



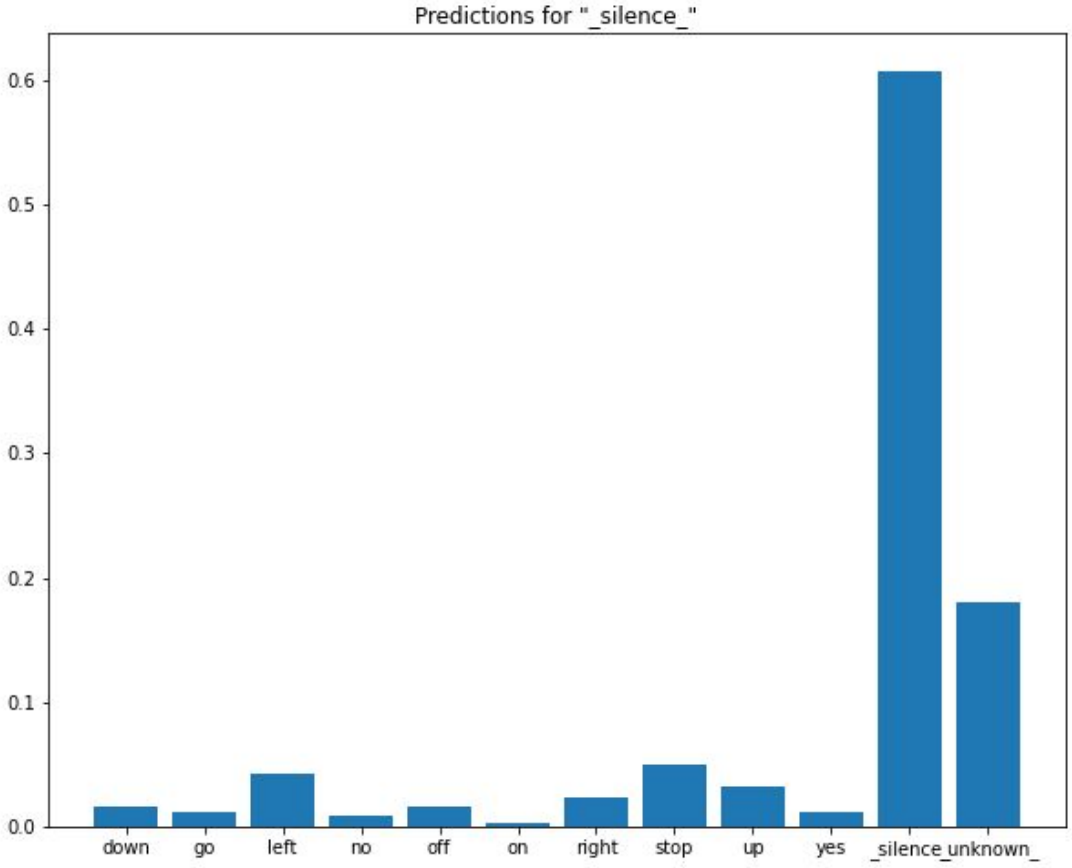
Predictions for “left”



Predictions for “up”



Predictions for “_silence_”



Streamlit App

Goal - load a pre-recorded speech command and predict the command spoken

Issue - preprocessing the new .wav audio file so that the model will be able to predict speech command.

Conclusions

CNN Model

- Success

Next Steps

- Expand Vocabulary
- Revisit the Modeling

Streamlit App

- Address preprocessing issue
- Record Sound Wave
- Add newly recorded data to model.

Thanks

Do you have any questions?