Summary Of Case-Study

- In this case study we have to predict word from audio signal.
- Total we have 30 unique word and total 65k data points.
- Most Audio Clips are 1 sec long and frequency of all clips are 16k.

Data Preprocessing

- For data preparation first I do padding for all clips.
- For padding we set one threshold we length of signal is < threshold than we add zero at the end of signal and if length of signal is > threshold than we truncate the signal.

Feature Extraction

- spectrogram : https://librosa.org/doc/main/generated/librosa.feature.melspectrogram.html (https://librosa.org/doc/main/generated/librosa.feature.melspectrogram.html)
- MFCC : https://librosa.org/doc/main/generated/librosa.feature.mfcc.html (https://librosa.org/doc/main/generated/librosa.feature.mfcc.html)
- lagfbank : https://pypi.org/project/python_speech_features/0.4/ (https://pypi.org/project/python_speech_features/0.4/)
- For Feature Extraction i use melspectrogram, MFCC, logfbank and Data Augmentation.

Modeling

Model 1

- In model 1 i use very simple sequential structure like Dense-->Activation-->Dropout.
- I got 69% train and 70% validation accuracy at 40 epoch. After 20 epoch model not learn much because it is very simple model.

Model 2

- In model 2 i use conv1d with BatchNormalization and LSTM layers.
- In this model i got 85% train and 80% validation accuracy at 10 epoch. After 7th epoch model start overfitting and 10th epoch i get 80% validation accuracy.
- Accuracy Improved from model-1.

model 3

- For Same architecture i use augmentation technique. In augmentation i do pitch shifting and time stretching. Because of less computation power i use only 200 points per categories(total 1600 points) for augmentation.
- After that total we have 1600 augmentated data points + 65k original data points.
- In this method i get 88% train and 85% validation accuracy at 20 epochs.
- Accuracy is Improved.

Model 4

- For Same architecture i use MFCC feature extraction method.
- In this model i got 90% train and 88% validation accuracy at 30 epochs.

Model 5

- In this model i use GRU and for feature extraction i use logfbank.
- In this model i got 85% train and 82% validation accuracy at 20 epochs.
- Accuracy droped.

Model 6

- In this model i use conv2d and for feature extraction i use MFCC.
- In this model i got 92% train and 88% validation accuracy.
- Accuracy is not improved.

Model 7

- In this model i use conv2d and for feature extraction i use log-spectrogram.
- In this model i got 97% train and 95% validation accuracy at 10 epoch.
- Accuracy increase drastically by using con2d and log-spectrogram.

Error Analysis- Reasons why error happen in model

1. All misclassification points have very low amplitude.
2. Amplitude means : The amplitude of a wave is related to the amount of energy it carries. A low amplitude means wave of signal carries a small amount of energy.
3. Many True label are also labeled incorrectly. Ex: 3rd data point is labeled as marvie(Ture label) and predicted as nine and it is actual nine(concluded by listen). **Here many ground truth labels are wrong.** (cause of error : human error)
4. Many points are silent but labeled as it is not silent.
5. As we see in count plot highly misclassified label is three. and very low misclassified label is yes.
6. From confusion matrix we say that model is not able to classify tree as tree and three as three.
7. we can see in confusion matrix highest misclassification happen in three and tree(both are very similar in pronunciation).
8. From Classification Report we can conclude that,

- **Model Perform Best**(where F1 score >= 95) in this categories : [bird, cat, four, happy, house, marvin, one, sheila, two, wow, yes, zero]
- **Model Perform Medium**(where F1 score >= 90 and < 95) in this categories : [bed, dog, down, eight, five, left, nine, no, off, on, right, seven, six, stop]
- **Model Perform Worst**(maximum misclassify, where F1 score < 91) in this categories : [go, three, tree, up]