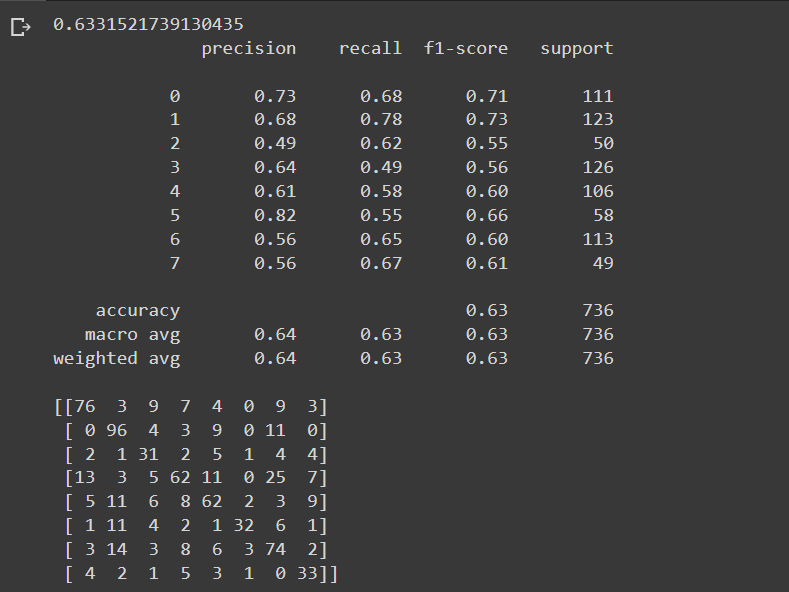
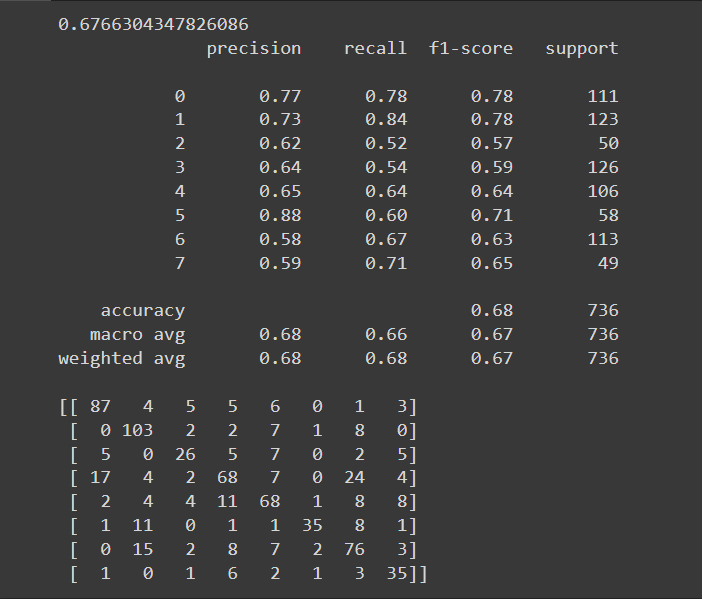
**Week 3**

* Initialised a random forest classifier with number of estimators =100 and the max\_features parameter set to none (max\_features=num\_features)

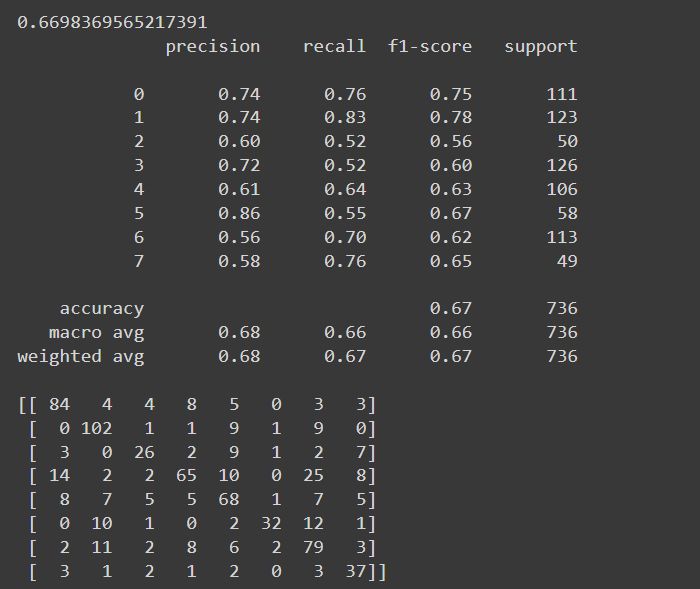


* Experimented with other values for max\_features:

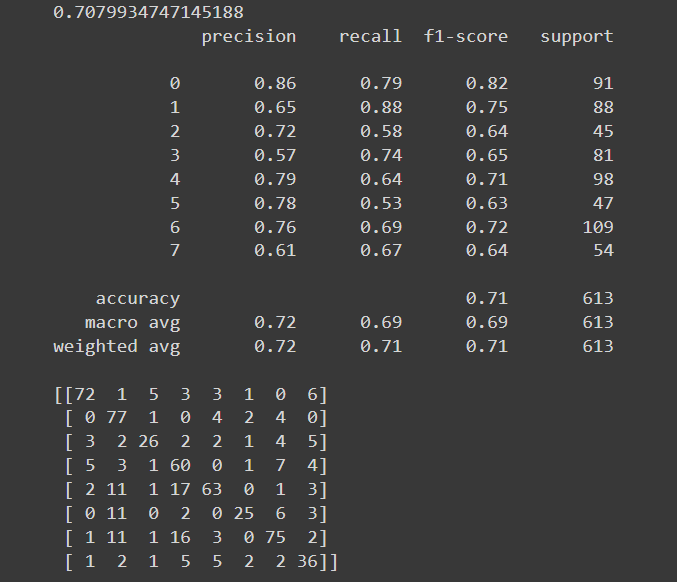
Sqrt:



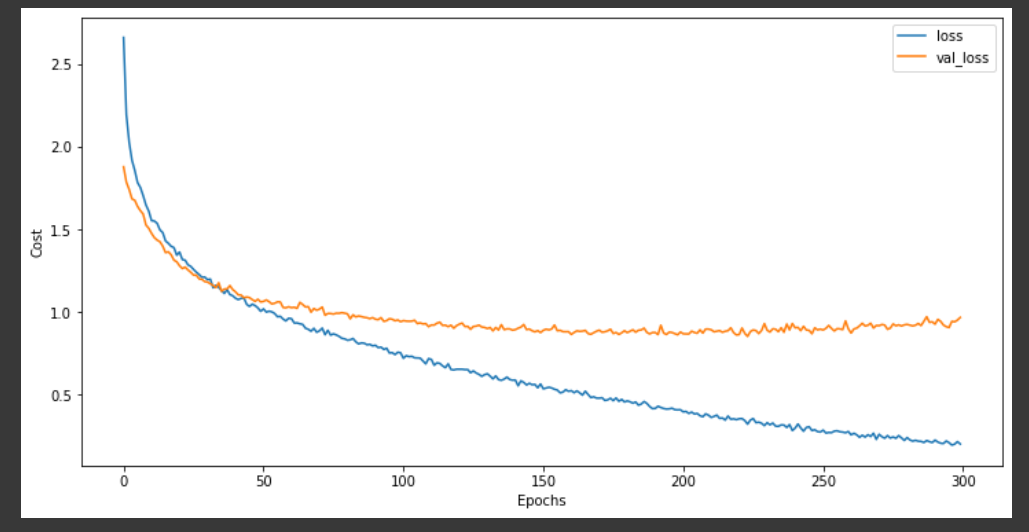
Log2:

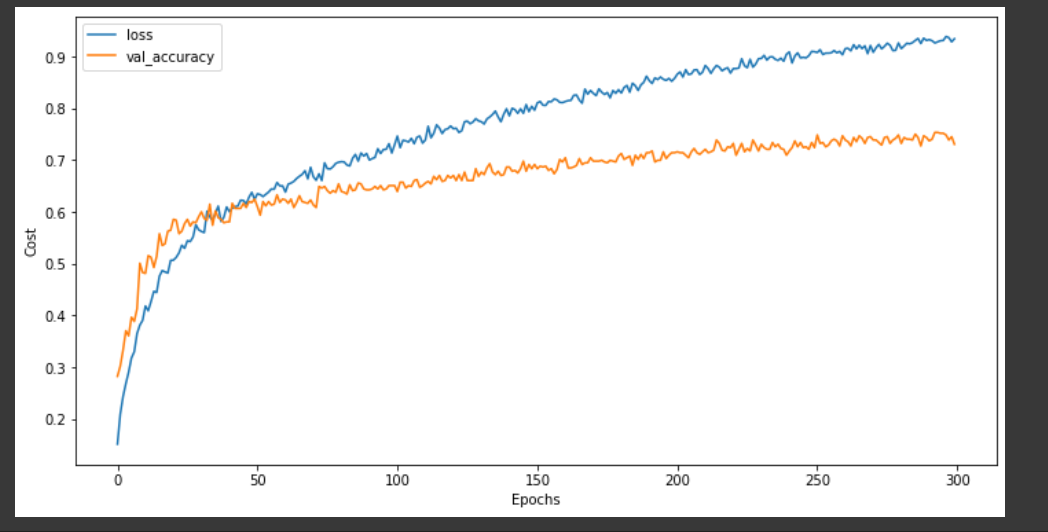


* Experimented with different values for number of estimators and got an accuracy of 70.1% with the following parameters: n\_estimators=250,max\_features=sqrt. (also changed the train test split to a 80 20 split)

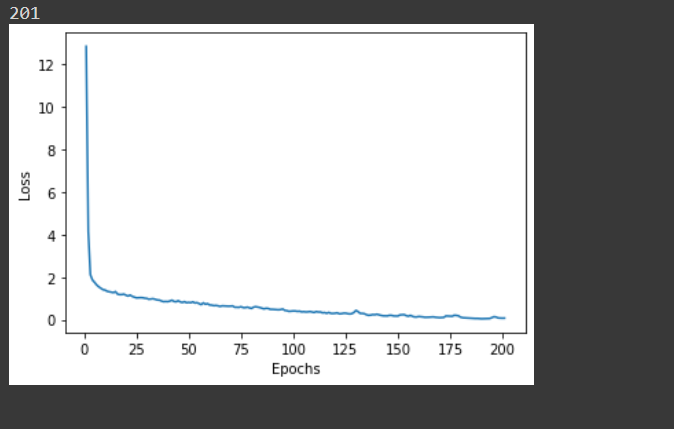


* Tested the accuracy of existing CNN model with the 80 20 train test split and got an accuracy of 73.08%.

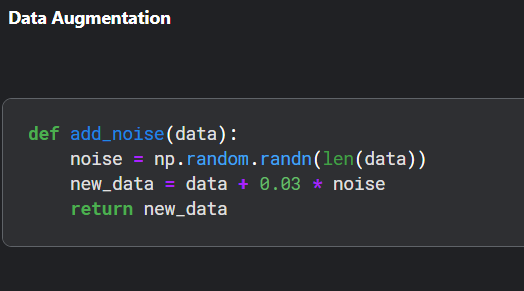




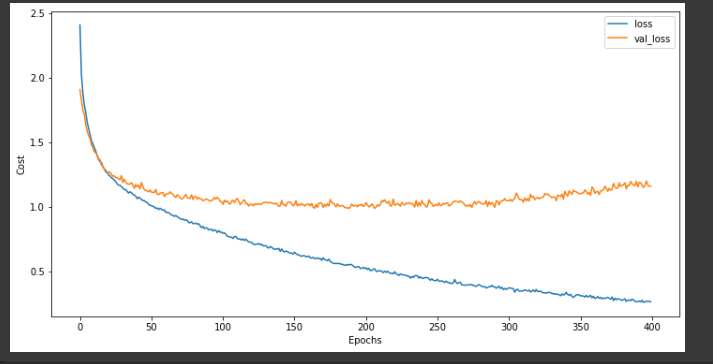
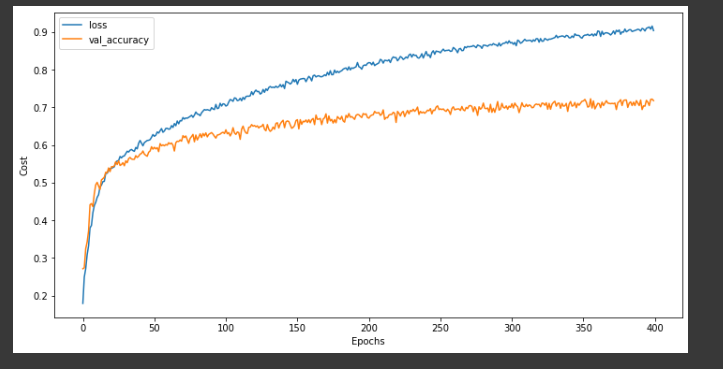
* Built a simple MLP model , experimented with the hyper parameters like the loss function(adam,sgd), the number of hidden layers and the number of neurons in the hidden layer, learning rate , regularization parameters to prevent overfitting.
* Built a model with 3 hidden layers with 200 neurons in each and 1 input layer with 195 neurons and 1 output layer with 8 neurons and got an accuracy of 73.4%
* Plotted the loss curve for the mlp to check for overfitting



* Brought in another dataset to increase the number of training samples to help increase the validation accuracy of the model.
* Appended data from the CREMA-D dataset which consists of 7,442 original clips from 91 actors. These clips were from 48 male and 43 female actors between the ages of 20 and 74 . The actors spoke from a selection of 12 sentences. The sentences were presented using one of six different emotions (Anger, Disgust, Fear, Happy, Neutral, and Sad) and four different emotion levels (Low, Medium, High, and Unspecified).
* One problem with this concatenation was that the number of data samples for the emotions calm and surprised were less and thus the accuracy of the models (as seen below) also wasn’t high.
* Tested the new dataset with the existing CNN architecture which produced a validation accuracy of 50 % , SVM which produced an accuracy of 51% and the Random Forest model produced an accuracy of 52%.
* Modified the architectures and tested the models but there was no significant improvement in the accuracy.
* Then tested the CREMA-D dataset alone with a CNN,SVM and a Random Forest classifier. Accuracies produced were 49.6%,43.1%,46.6% respectively. This was due to the fact that the dataset contained a lot of noise .
* Data Augmentation was performed on the original dataset where a random noise was added to the existing data to provide more data to help increase the accuracy of the model.



* Trained the existing models with augmented data . There was no significant improvement in the accuracy of the model but the training process was more controlled and the over fitting problem was eliminated to an extent.



* The emotions/ classes predicted are 'surprised', 'neutral', 'disgust', 'fear', 'sad', 'calm', 'happy','angry'. However prediction of emotions like surprised , disgust maybe trivial for our use case(detection of emotions in a vehicle) and thus removal of data containing these emotion classes may help increase the accuracy of the models.
* The emotion classes ['angry', 'sad' , 'neutral' , 'happy'] were retained and the others were discarded. Data augmentation was done and then these data samples were used to train CNN , SVM , Random Forest and MLP classifiers .
* Accuracies of 84.8% , 72% , 81% and 86% were produced for CNN , SVM , Random Forest and MLP respectively.
* Pre-processing of spectrogram images to tackle the Speech emotion Recognition using an alternate approach was done.

**Tasks for next week:**

* Speech emotion recognition using the spectrogram images.
* Testing the models with other datasets and real life data.
* Finalising models based on their performance on real life data and use cases.