# Implementation

**Choice Of Dataset**

Our project focuses on getting data from the audio dataset so that we can implement Content based Filtering. In this project, our choice of Dataset is GTZan Dataset. GTZan Dataset contains 1000 songs of 30 seconds which is equally divided in to 10 genres. **Fill more**

**Generating Mel Spectogram**

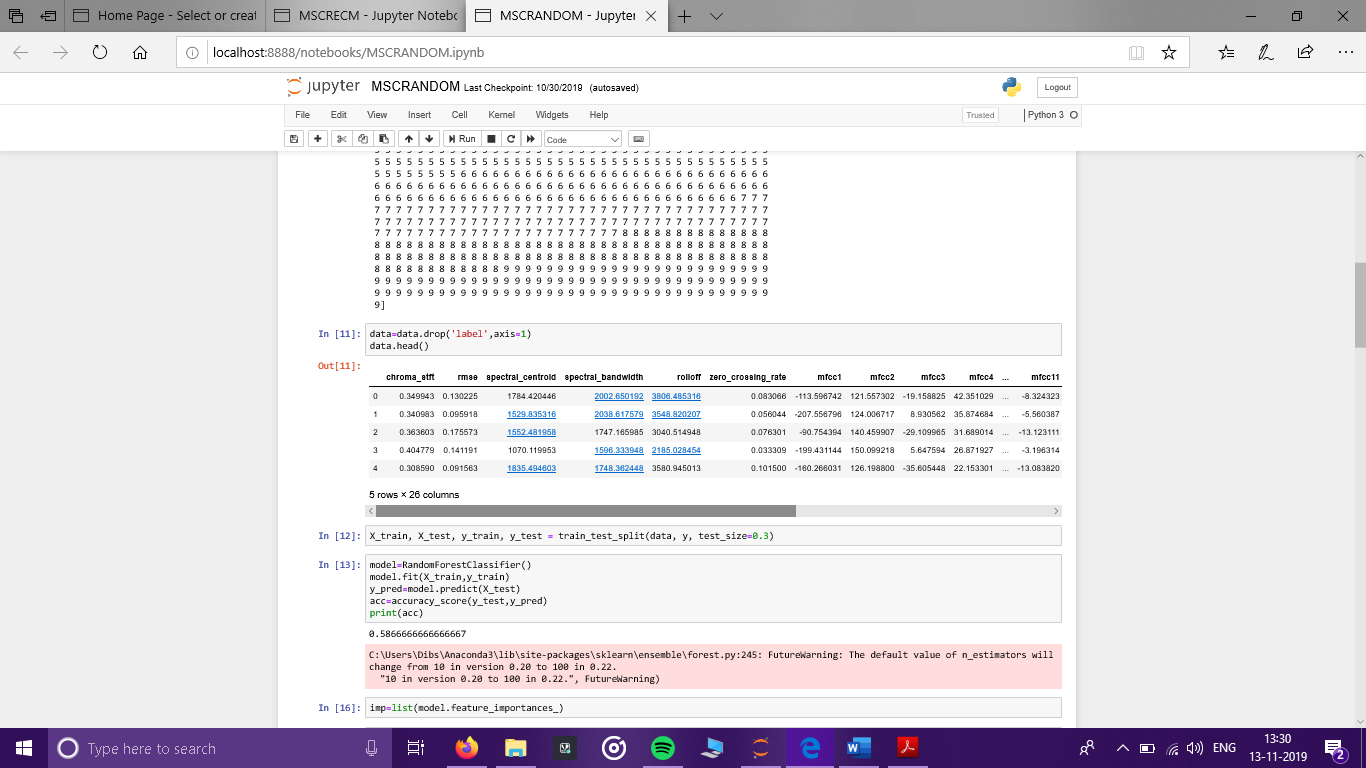
Each one of audio files from the dataset was taken and converted into equivalent spectograms. Spectogram can be defined as a visual representation of the complete spectrum of all the frequencies with respect to time. It can be seen as squared magnitude of short term Fourier Transform(STFT) of the input audio signal. A Fourier transform is basically a representation that inputs a signal in time domain and outputs in frequencies. Mel Scale Spectogram uses Mel Scale in the y-axis to get Frequencies in different bins. This Mel scale Spectogram generation requires librosa built in library which directly converts an audio input into this. The function takes some parameters like – window length (window of time to perform transform) and hop length (number of samples between successive frames). We have used a window length of 2048 (appx 10ms). Hop length was taken as 512. Using Mel scale, we can distinguish between audio better as it scales the Hz scale using log function into dB(Decibels) and squashes the difference to relate more to human perceived pitch.

|  |  |
| --- | --- |
| C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\CBC96D91.tmp | C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\EA2AD6C7.tmp |
| C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\D818B36D.tmp | C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\93BB0503.tmp |
| C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\1A04C909.tmp | C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\AE7958FF.tmp |
| C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\2AF8CA65.tmp | C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\202B4EBB.tmp |
| C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\D2E39381.tmp | C:\Users\Dibs\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\DC902237.tmp |

*Figure: Spectogram plot (From top left): Rock, Classical, Country, Blues, Disco, HipHop, Jazz, Metal, Pop, Reggae*

**Model and Approach**

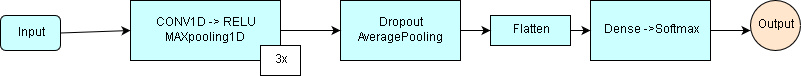
Using the MFCC features , We already created our feature dataset. It consists of RMSE, Spectral Centroid , Spectral Bandwidth, Roll off , Zero crossing rate and the other 20 MFCC Features. These completely makes our Feature dataset.

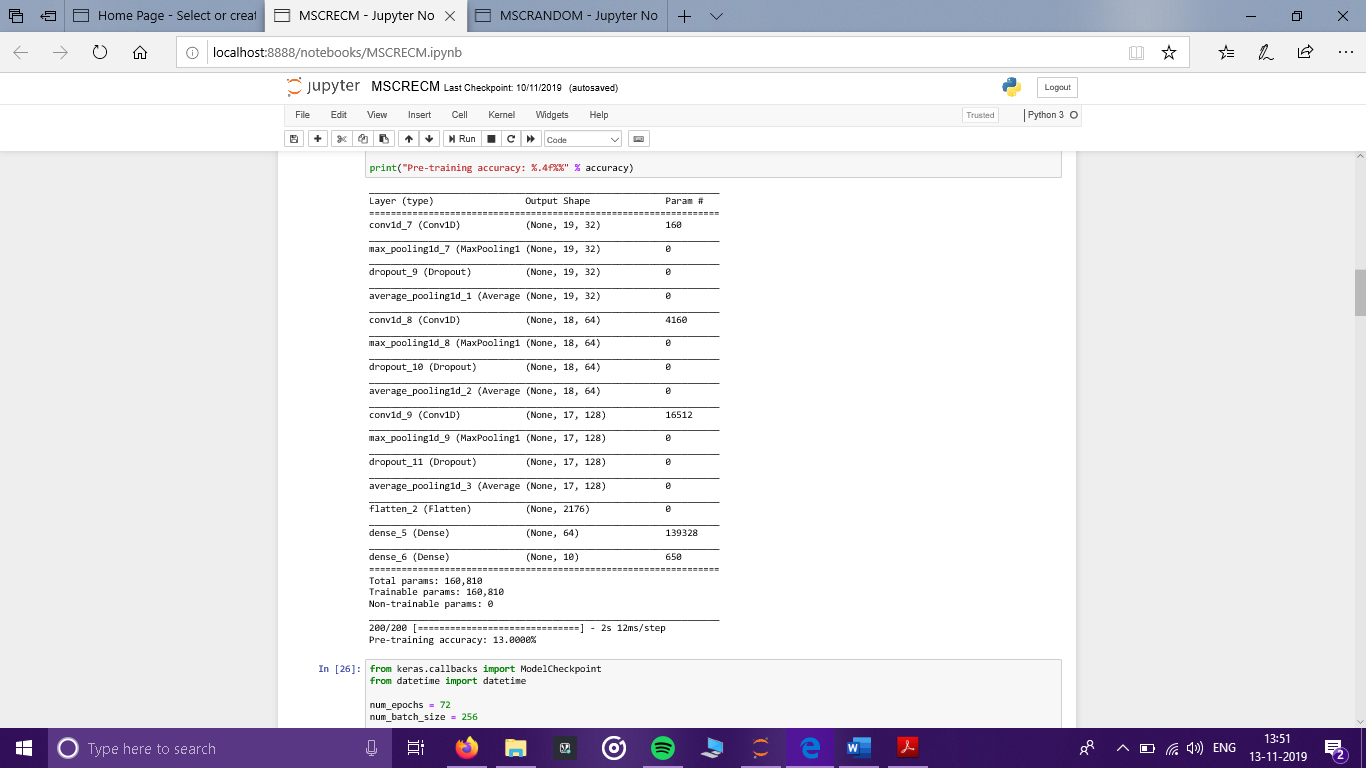


*Figure: Feature Dataset extracted from our GTZAN songs.*

**CNN Model**

Image recognition related tasks have been using Convolutional Neural Networks (CNN’s) since a long time. They work on the simple logic of Convolution over the complete data to understand the underneath layout. Since CNN’s have a good precision on Image dataset, we thought to use this on our Mel Scale Spectograms to get better output results.



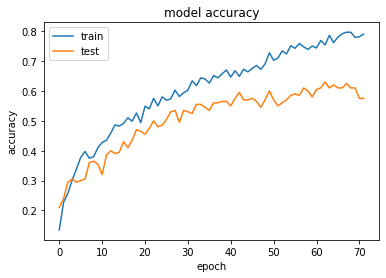


*Figure: Model information by Keras.*

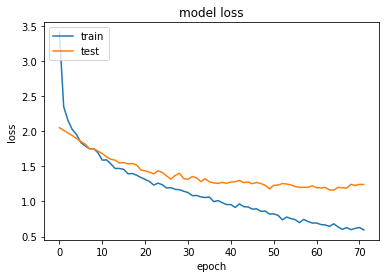
The model was trained with Adam optimizer and loss function categorical cross entropy. The model was then trained for around 72 epochs with batch sizes of 256 used.

After the Hyper parameter Tuning We were able to Get Accuracy of 92 % on Training and 54 % on Test Dataset which is much higher than a regular CRNN network model as well.

We then plotted our training and validation Accuracy and loss Graphs to check out our work.



*Figure: Accuracy over the Epochs*



*Figure: Model Loss over the Epochs*