Usecase 3 :

Million songs : Year prediction dataset

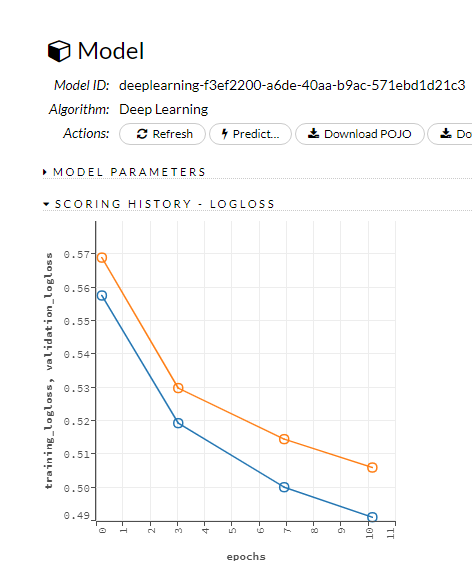
The dataset contains a the features of a million songs from the year 1922 to 2011.

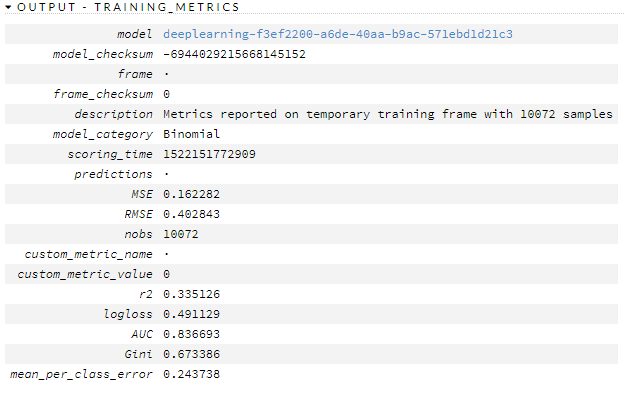
Use case is to predict whether the song is released before 2004 or after 2004, taking into consideration the music features.

The data set is classified binomially, 1 and 0 stating whether its before 2004 or after 2004.

I have used deep learning to predict the year of the songs taking the timbre features of each song.

Scoring history :



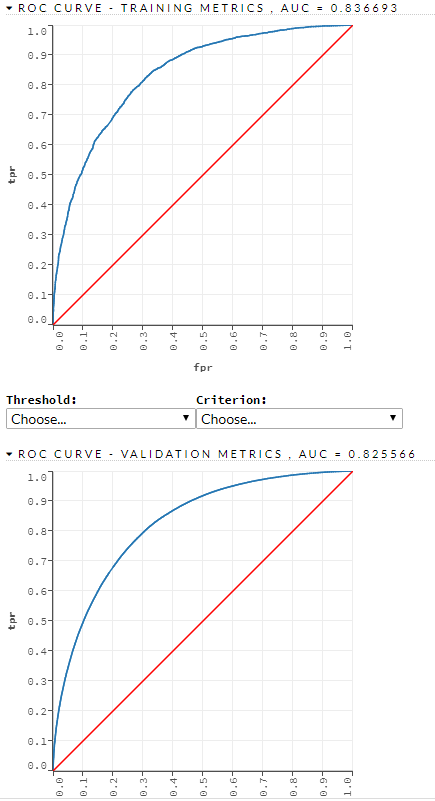


The scoring history shows the error rate has improved by 6% approximately ,I.e., 55% decreased to 49%

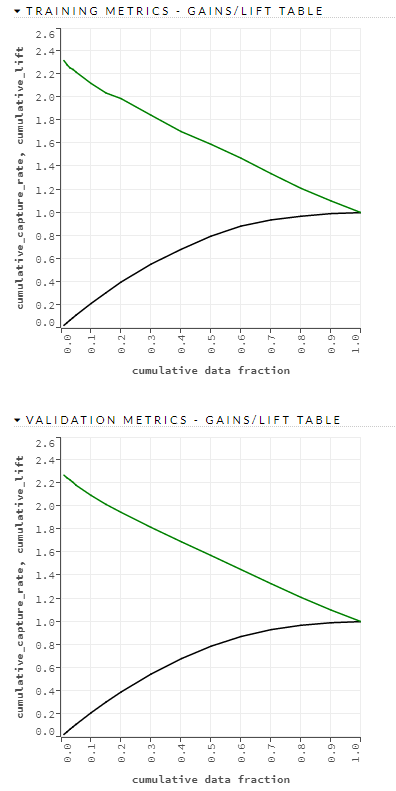
The R2 is 33.5%. It tells that the data can be improved by 33.5%. We can get even better results if the R2 is higher. Hence, the model can be improved.

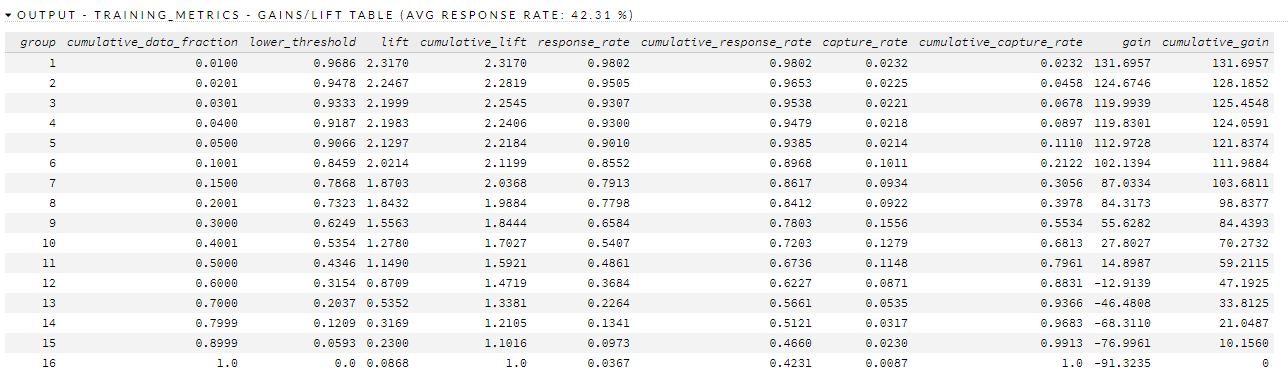
MSE and RMSE ,I.e., the mean square error of the data is low, which shows the prediction has less error.

We have a really good AUC,I.e., 83.6% ,which explains the model is highly predictable.

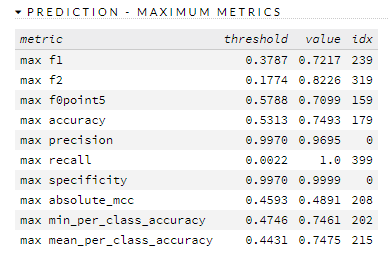


The AUC of Training and validation set is close, which shows our model is not overfitted/underfitted.





In the above figure , it illustrates the lift in the response of the training metrics is 42.31% . In groups 1 to 11, we observe the lift is good , which says they are more likely to be after 2004.



The model gives us the overall statistics which says the maximum accuracy is 74.9% at threshold 53.13% which is a good value. It states that the predicted result is 74.9% accurate.

No. Of true negatives(Sensitivity) is 0 for 0.99% which shows the model is highly accurate.