

Using Linear Regression to Predict the Streams of a Song on Spotify MVP

Spotify is a digital music, podcast, and video service that gives you access to millions of songs and other content and currently it has around 365 million active users. This project aims to give a linear regression model to predict the number of streams of the top 200 songs on Spotify each year from 2016-2021 based on many features the song has.

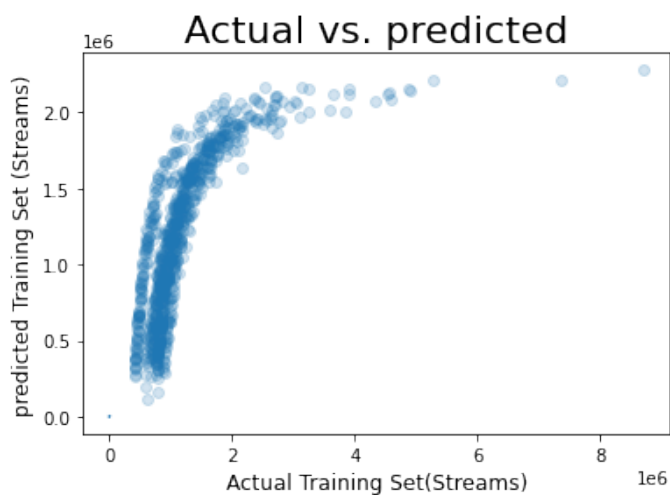


Fig a. Actual Vs Predicted values of the baseline model

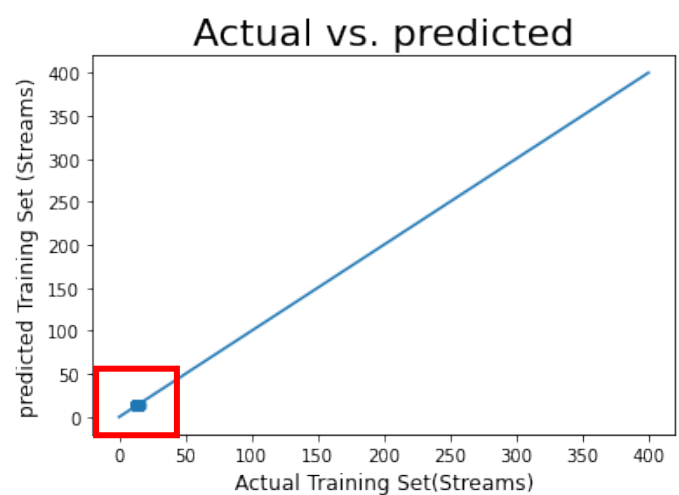


Fig b. Actual Vs Predicted values after improving the model

To achieve this goal, A linear regression model was built, a scatterplot of the Actual data VS the predicted data is shown above. To give better prediction and higher accuracy many approaches has been taken in consider as you may see in Fig b, the model has been improved such that the error shown in the red square is way less than the model in Fig a.