G18

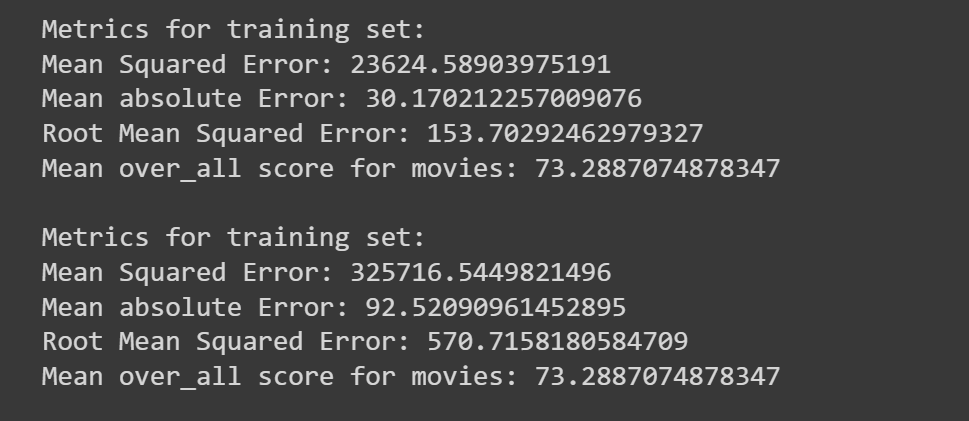
Performance Metrics for ML Model

This document shows the results we got for training our model with different variations.

We were primarily trying to predict “success\_overall” and had this set as the target for our ML model.

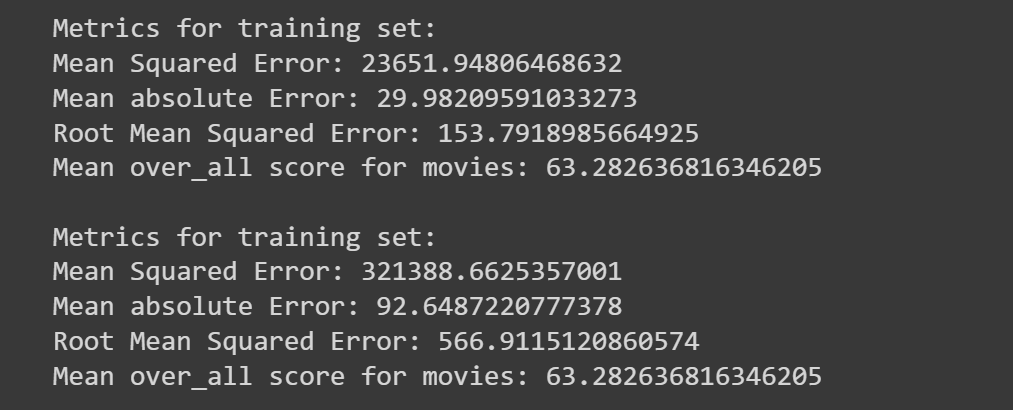
We started off with using bagging, with the following parameters in our dataframe: Budget, Year, Month, Directors, Production House, Genre (the last three of which were one-hot encoded)

The results were as follows:



We felt the model was overfitting, and so we tried to fix this.

Results after using z-score normalizing on Budget:



We tried to simplify the model by removing the one-hot encoded ‘Production House’ feature:

A computer screen shot of a program

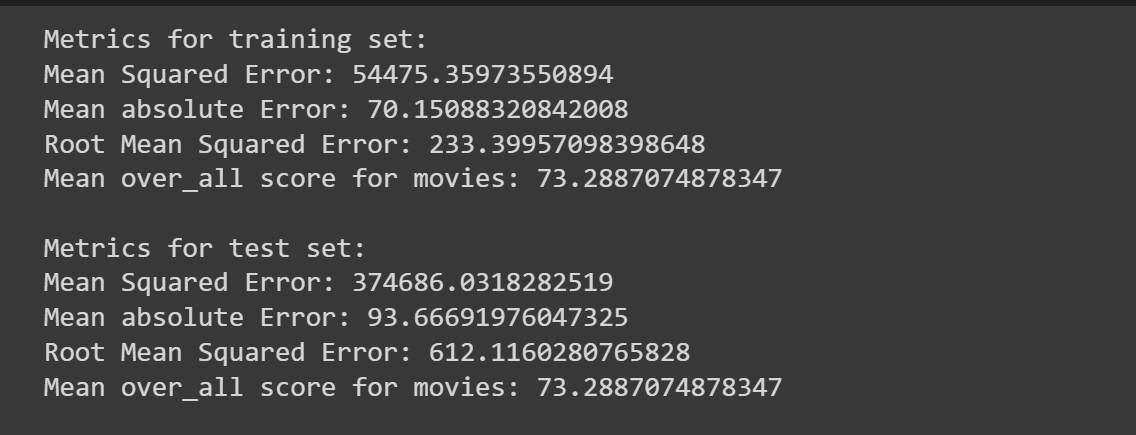
Description automatically generated

We felt it could be beneficial to one-hot encode Month as certain months may have higher movie success due to seasonality, but they are not linear or linked (as in, month 5 and month 6 may not connected in terms of seasonality):

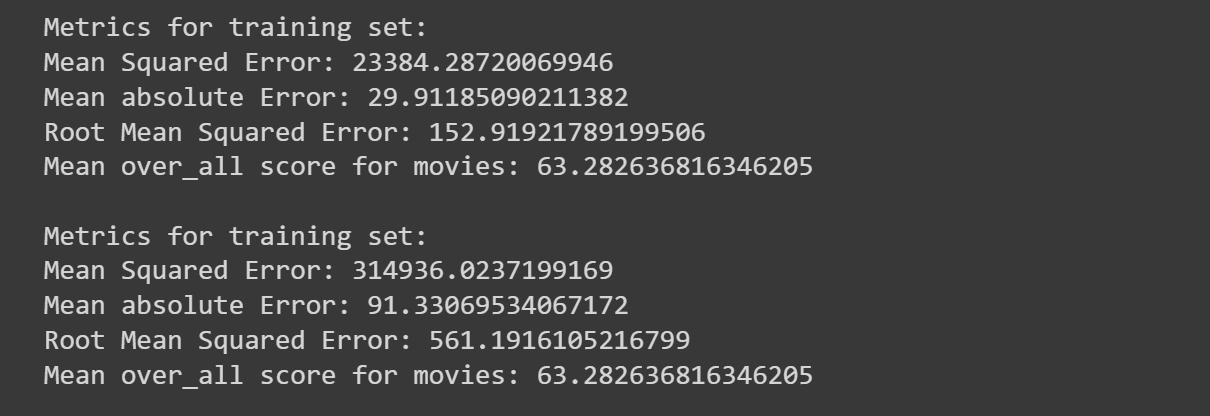
A computer screen shot of a program

Description automatically generated

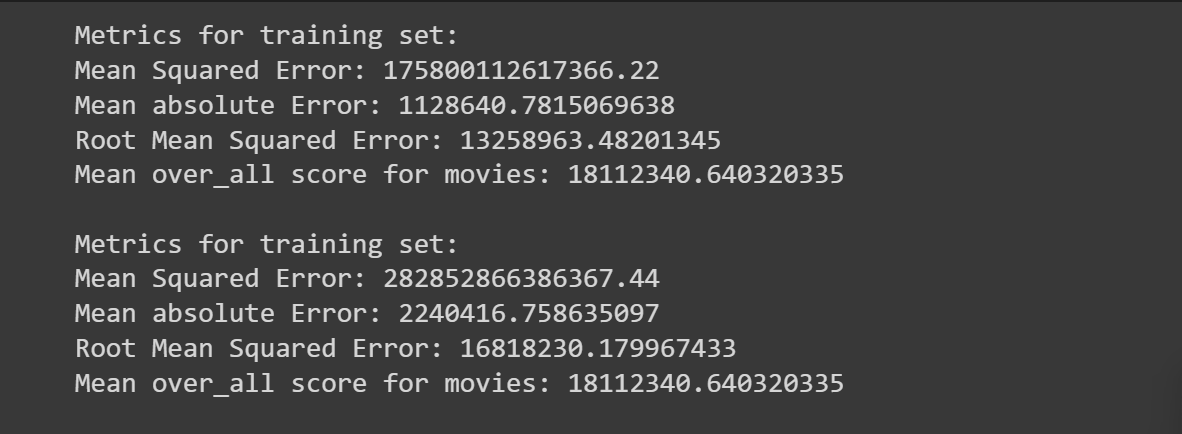
We tried to use boosting instead of bagging:



We tried to predict just success\_financial (instead of success\_overall):



We tried to predict Box Office (instead of success\_overall or success\_financial):



Unfortunately, despite all the attempts, the model is not performing as well as we had hoped. We feel this could be due to two primary reasons:

1. There may be too few movies per director (around 20 on average) and that may not be enough data for the ML model to train, when there are 242 directors in the dataset that it has to learn the patterns for. The model may not have identified that the one-hot encoded columns for the directors are very important.
2. When we split the dataset into train and test sets, we do it randomly, which should be fine, but it may result in some directors getting most of their movies in just the test set and not the train set. This would result in the model not learning about them.