06/14/2020

Group Proposal

We, group 1, selected a data set including different information movies for our project. We have selected this dataset because it is not only up to date but also, it has a wide variety of movies with different categories. We are going to use these movie's categories such as runtime, number of votes, rating, and metascore to predict the <u>revenue</u> of these movies. Through analyzing the impact of different factors on the revenue of movies, we can give some suggestions for movie companies to make what kind of movie in order to obtain high revenue in the following year. The data set has 1000 rows which makes it large enough to train a machine learning model or network. We are going to use MLPRegressor from the neural network in sklearn. Depending on the dataset, we might need to alter the network. We are going to use PyCharm as our software because it has many built in networks as well as it is assigned by our class. Our reference material is mainly based on our textbook *Neural Network Design*, and we will use Github as one of our reference sources, when we have some coding issues in the process of our project. We are going to use the mean square error or R^2 score to judge the performance of our network. We are going to clean our data for the first couple of days. Then, we are going to build our model and lastly analyze our model based on our test scores and metrics.