Procedures

The neural network was built and initialized in Python. Python 3.7.3 was the latest version used for the experiment. The libraries installed and imported into python include numpy, sci-kit, pandas, seaborn, and matplotlib. The experiment will function properly on any IDE.

The data for the experiment was collected from many different sources including the-numbers.com, rottentomatoes.com, boxofficereport.com, and screenrant.com. The four subjects of data collected were the movie title and year, production budgets, the amount of participating theaters on opening weekend, amount of views for the movie’s trailers, Rotten Tomatoes score, and whether the movie succeeded or not on the first box office weekend. A Rotten Tomatoes score is a percentage from the website rottentomatoes.com which is the average score of all critics’ ratings of the movie and can heavily affect the number of people that see the movie on the opening weekend. All of the data was numerical and mostly large numbers in the millions, which was scaled down to hundreds using units in millions. The data that has been collected can now be transferred into a CSV and used in the neural network using pandas to read the file.

The data was then shuffled for the neural network. 60-80% of the data was used to train the neural network to increase accuracy, while the others were used for testing.

Finally, the program was tested and evaluated. Using sklearn the layers were initialized for the neural network and then tested to see how accurate the program was.