**Film Jamz**

**Support File, Part 1**

Jingyi, Aidan, Zhimin, Michael

Before implementing our desired algorithm, random forest regression, we preprocessed our data which contains 6820 records. We used 10 predictor variables to decide what is the best feature when predicting how well a movie will do. The variables are company, country, director, genre, gross revenue, name, rating, runtime, score and votes. After completing the preprocessing, we applied random forest regression algorithm, which is one of the most effective machine learning models for predictive analytics. The random forest regression algorithm is a classification algorithm. What it does is create a forest with several trees. The more trees, the more robust it will be. In the same way in the random forest classifier, the higher the number of trees, the higher the accuracy results. Once we ran the algorithm, it was revealed that ***rating*** was the best feature when deciding how well a movie will perform. We then proceeded to graph our data where it showed rating as having a remarkable weight advantage versus its counterparts.