[](https://www.iss.nus.edu.sg/)

**Machine Learning**

***Project Proposal***

**Submitted By: Team 7**

|  |
| --- |
| **Student Name** |
| Aung Myo |
| Lim, Hean Soon |
| Lin, Wei |
| Liu, Yakun |
| Nyi Nyi Zin |
| Cheng Hao |

1. **Business Problem Formulation:**
2. To predict whether a certain criteria/combination of Actor(s) and Genre will be a hit or a miss. This will help a movie producer to determine whether the proposed movie is worth a shot.
3. **Objective:** Build a Machine Learning model to predict whether a potential movie with a certain actor on a certain genre will be a success.
4. **Machine Learning Approach (Learning Algorithm):** Classification and Decision Tree
5. **Programming Languages to be Used:**

Java with Weka

1. **Model testing:** Data from IMDB
2. **Model deployment:** As we feed in different sets of data, the machine should learn the success rate of the genre paired with the actors with the current age. This will tell if the proposed movie will be a success or not.

In order to increase the accuracy of the learning algorithm, we will be feeding in over 4000 data sets of movie and actors information [1]

The dataset information will contain information such as:

* Main Actor popularity
* Secondary Actor popularity
* Director popularity
* Genre
* IMDB score
* Country of Origin (of the movie)
* Gross Profit
* Budget

From the dataset, the machine will classify a movie to be a success if the gross profit is 65% more than budget.

Next, from the classification, the machine will identify the success factors by:

1. Main actor’s popularity
2. Secondary actor’s popularity
3. Combination of genre (more than 1 genre will take into count)
4. Director

Through the 3 items above, the machine will be able to predict whether the new movie will fall under the success classification or failure

We believe the above criteria is the main attraction for moviegoers.

**Data Preparation and Extraction:**

1. Select data from reliable data source
   1. IMDB
   2. Facebook
2. Preprocess the extracted (clean, format)
3. Transform (decomposition, aggregation)

Reference:

[1] [http://www.imdb.com/interfaces](http://www.google.com/url?q=http%3A%2F%2Fwww.imdb.com%2Finterfaces&sa=D&sntz=1&usg=AFQjCNG2S2dTotxBdZfofpEBBjfYsAOwGw)  
[2] [https://www.facebook.com](https://www.google.com/url?q=https%3A%2F%2Fwww.facebook.com&sa=D&sntz=1&usg=AFQjCNGvrk6qg81TNhGglNV3ezjav6AxdA)

[3] [https://www.kaggle.com](https://www.kaggle.com/)