**Predicting the Success Rate of Indian Movies**

**OVERVIEW:**

Bollywood is one of the biggest movie industries in the Asian continent. It produces around 150-200 movies every year. It has generated 3 billion USD revenue in the year 2011. It is also estimated that, it might enhance its revenue to 4.5 billion USD by the year 2016. But the question is, whether it is going to be easy task for it to reach the target.

Yashraj movies are one of the biggest production houses in India which produces 30 films every year. But not all movies they had produced were successful. We would like to use the data driven approach to estimate their success/failure rates. The approach can be used to help Yashraj to produce more successful movies and increase their revenue.

**Below is a one of those examples through which we are motivated:**

• Amazon.com have used data mining approach to predict the type of show that would be liked by viewers

• They had 14 pilot episodes posted on their website

• They monitored the number of viewers watched, comments and patterns on their site

• They also captured the rating for those pilot episodes, how much it was viewed and how many shared it with others.

• Based on the prediction they released their first TV show “Alpha house” which received 7.5 rating on IMDB.

**Data mining problem:**

The project begins with the unsupervised data model and proceeds with the generation of different data model techniques.

The some of the data mining problem that we might encounter are as follows:

Selection of attributes those are relevant to get the target variable

Selection of the number of years to predict target variable is also difficult

**Target variable:**

• Using the data sets available we are trying to predict what kind of movie would be successful

• This will be an unsupervised approach as we are using classification model to predict the target variable

**Features useful:**

The useful features can be such as:

1. Year and Month of the movie release.

2. Actor

3. Director

4. Genre

**Business value:**

The prediction would help the production house to gain more profits and avoid the possible errors.

**DATA:**

**Data Instances/Attributes:**

We collected the data from the IMDB database which is an open source and will be sufficient to predict the success rate of the future productions from Yashraj productions using WEKA tool. All the data in the source is in text format and should be converted to executable format. The data attributes include the following:

• Number of Votes

• Genre

• Running time

• Languages

• Actors

• Directors

**PROJECT MANAGEMENT**

**Team members:**

1. **Mani Karthik Annapureddy – Project execution**
2. **Rajendra - Interpretation**
3. **Meghana Sadhula - Analysis**
4. **Srujan – Data Collection**