**Movie success predictor**

**Problem statement**:

Predicting whether a movie will be financially or critically successful based on previous datapoints like cast, director, date of release, budget, genre, run time etc.

**Prospective Clients:**

This project will help movie production houses on deciding whether or not to produce a movie. It might help cast members to make a decision on whether to pick up a particular role or not.

**Dataset**:

Either the movielens 20M database

<https://grouplens.org/datasets/movielens/20m/>

or imdb5000 movie database

<https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>

Movielens data is much bigger in size. It has 27000 movies and additional fields like user defined tags and user ratings. So it is a better database for recommendation engines as well.

**Approach**:

Step 1: Scrub the data to extract pertinent information.

Step 2: Find connections between different factors

Step 3: Use the data for training the algorithm

Step 4: Use it for prediction/evaluate the model

**Patent Litigation Dataset**

**Problem statement**:

Finding correlation between patent litigation and success/failure of companies, growth trends of the industry in question etc. Popular perception is that industries that have a lot of ongoing litigation are the most competitive so should have the most rapid growth. Analysis of this data will help uncover the truth behind this consideration.

**Dataset**:

[https://www.kaggle.com/uspto/patent-litigations](https://www.kaggle.com/uspto/patent-litigations%20)