# Mentor: Ikechukwu Okonkwo

# Mentee: Mousom Mondal

Capstone 1: Movie success predictor

Model a movie predictor for revenue, box office hit and user ratings based on variables like budget, genre, cast, cast rating, director, director rating, month of release, # of theatres released in, released in overseas before or later etc.

## User/Client

Generally the movie studios, producers and investors can make critical decision and change several parameters based on the predictor. The model will use historical dataset with varied weightage in decision making – higher weightage to the recent movie statistics

## Data

[IMDB](ftp://ftp.fu-berlin.de/pub/misc/movies/database/), [Boxoffice](http://www.the-numbers.com/movie/budgets/all), [Ratings dataset](http://grouplens.org/datasets/movielens/latest/), [Tweets Rating dataset](https://github.com/sidooms/MovieTweetings)

Approach

* Collect and merge all the data - this will require significant wrangling.
* Scope the problem down to make it both more tractable, as well as relevant.
* Model it as a regression problem to predict revenue (in $). Train the model on an earlier subset of a data set and test it on later subsets.
* Evaluate effectiveness of the model.

## Deliverables

Code, Report, Slide Deck

#Capstone 2

Demographics – education, race, geographic location, religion,

(In progress)

Capstone 3

(In progress)