

Project Luther:

Predicting Return on Investment of Movies

San Francisco, 10/07/2016, Herr Nils

Goal: Predict Return on investment before the movie is produced

1 Target

$$\text{Return on investment (RoI)} := \frac{\text{Worldwide adjusted Gross} - \text{adj. Budget}}{\text{Adjusted Budget}}$$

10 Features

Budget
(adjusted)

Runtime

Genre

Reputation of
Actors

Reputation of
Director

True story?

Book
Adaption?

MTAA Rating

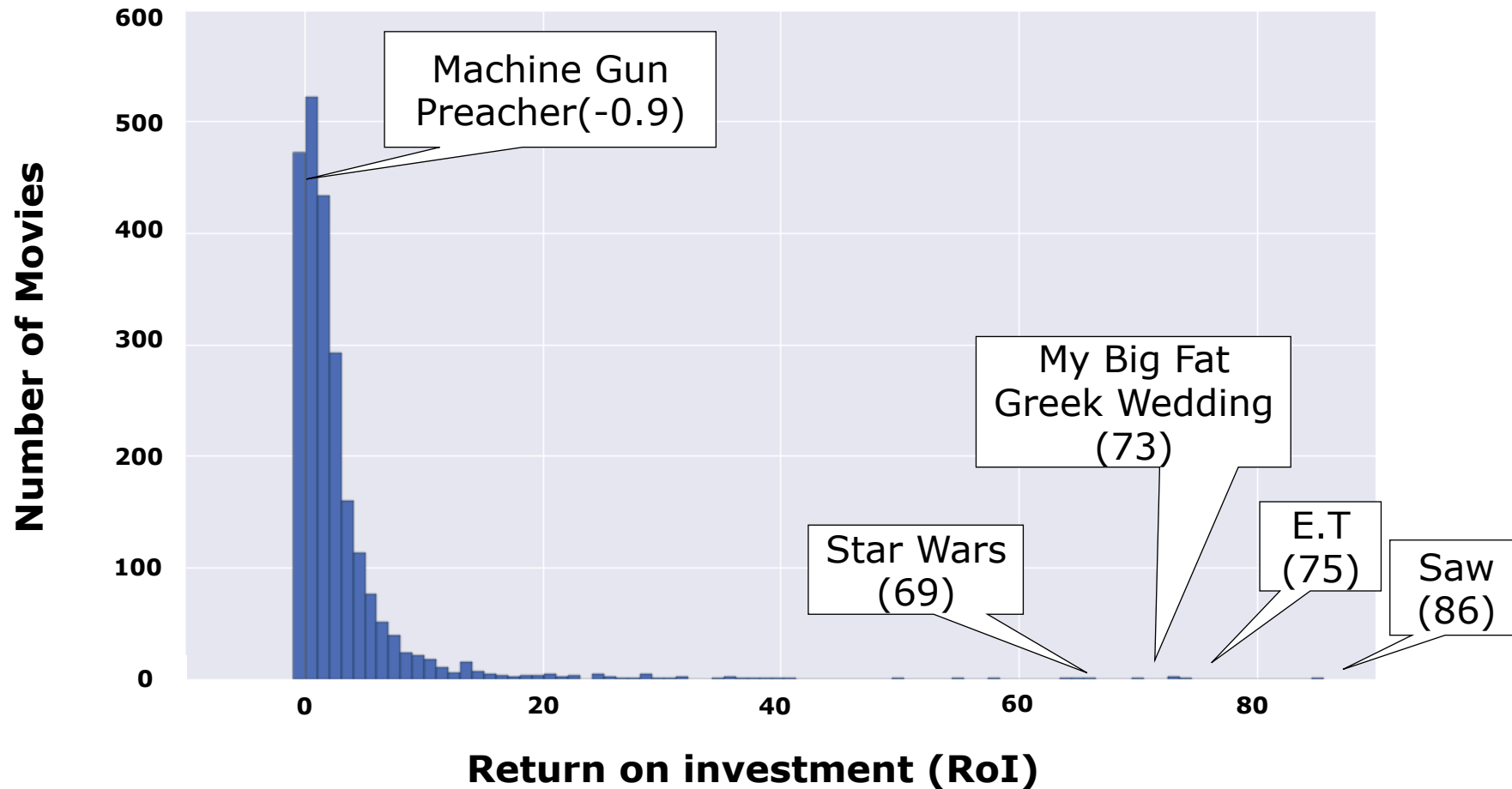
Season

3D enabled?

2400
Movies

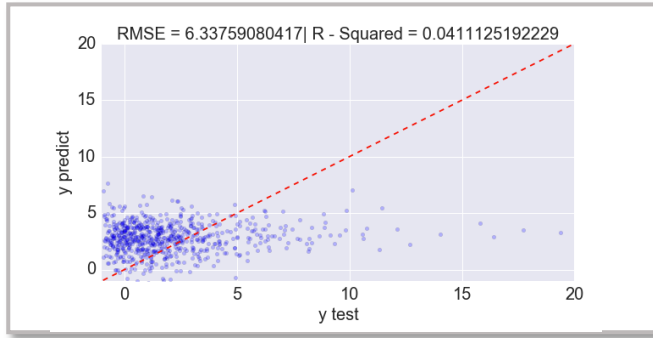
2400 Movies scraped using three sources:
Box Office Mojo, IMDb and goodread.com

Overview: RoI varies between -1 and ~84 between movies



Result: Random forest model with R^2 of 0.26 is more accurate than Linear Regressions

Standard Linear Regression

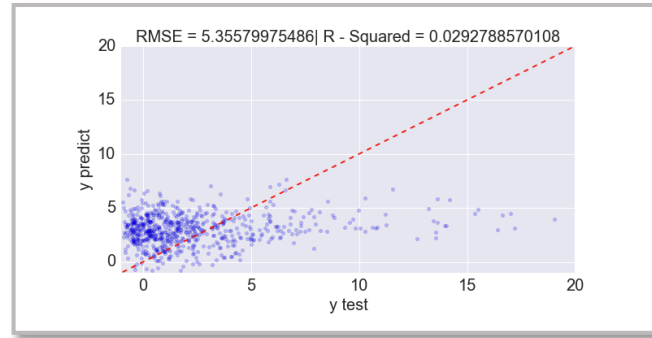


- **Highest correlations:**

- **Adj. Budget:** -0.14
- **Action/ Adventure:** -0.07
- **Rat-PG-14:** -0.6
- **Actors' Reputation:** 0.05

R^2 : 0.04

Lasso Regression

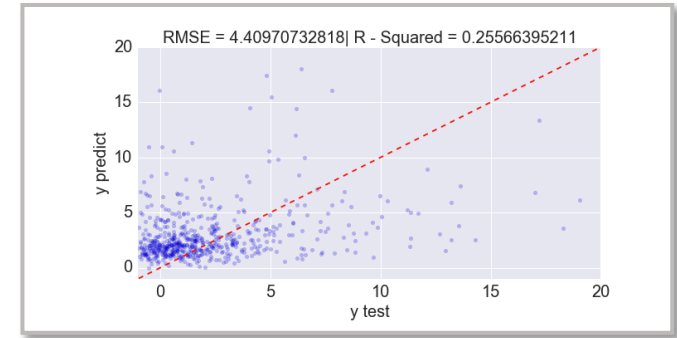


- **Highest correlations:**

- See Standard linear regression

R^2 : 0.03

Random Forrest



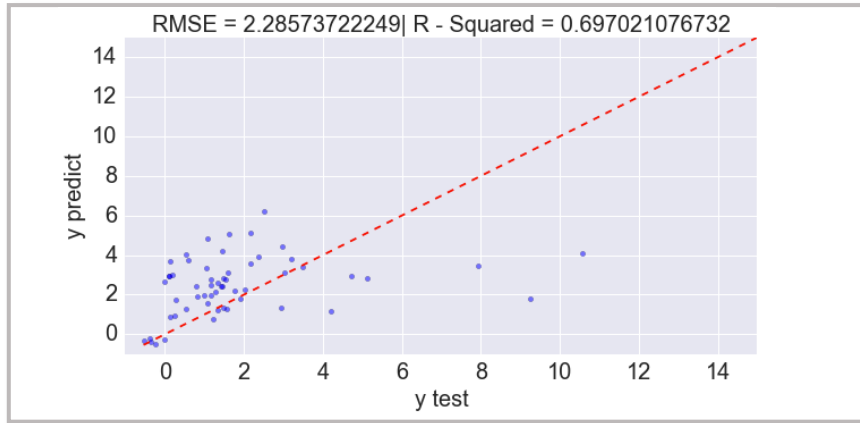
- **Most relevant features:**

- **Adjusted Budget**
- **Runtime**
- **Actors' Reputation**

R^2 : 0.26

Major accuracy improvement when looking within selected Genres only

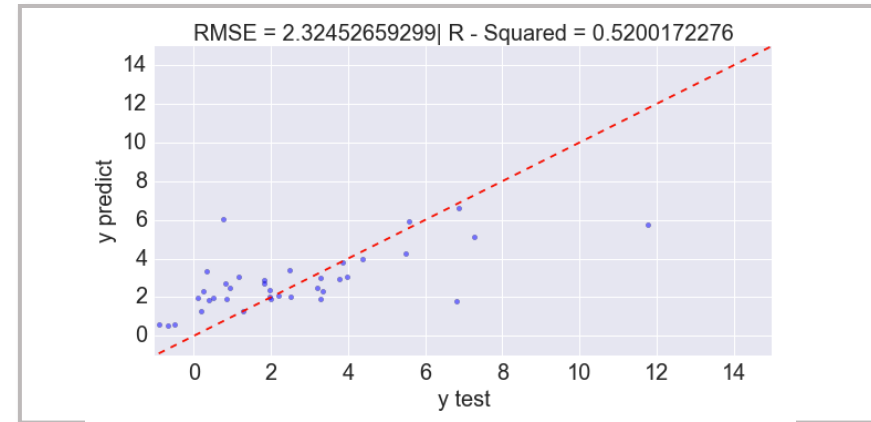
Sci-Fi/ Fantasy Movies only (Random Forrest)



- Random Forest Model for Sci-Fi/ Fantasy movies
- 228 movies in dataset

R^2 : 0.70

Animation Movies only (Random Forrest Model)



- Random Forest Model for Animation Movies
- 149 respective movies in dataset

R^2 : 0.52

Potential Next steps

- Examine additional subsets – e.g.
 - Different rating categories
 - 3D movies only
 - ...
- Consider further regression models