

1. Introduction

• Monetary motivation

- 500 movies/year produced in the U.S. on average
- \$60 Million of investment capital per film on average
- Success: profitability*

• Acquiring Investments

- Granted uncertainty, what to invest in
- Assure investors based on info available at the time*
- Many informing features unavailable

• Goal: Accurately predict success to aid in:

- Increasing investor certainty

2. Feature Engineering

• “What”

- Features related to movie content
- Basic: genre and rating
- Plot synopsis reflective of script:
 - Employ Latent Dirichlet Allocation

• “Who”

- The individuals responsible for portraying the movie
- Star power features (actors and director)
 - “Name brand” effect
- Team tenure
- Collaboration network-based features

• “When”

- Capture temporal ebbs and flows of the movie industry
- Also capture “what” is popular “when” – ie. Annual Weighted Profitability by Genre:

$$AWPG_m = \sum_{m'_i \in y-1} sim(g(m), g(m'_i)) * p(m'_i)$$

3. Profitability

• Defining success in terms of profit

- Revenue/Budget (profit ratio)
- A two class problem (profitable vs. not profitable)
 - Top 1/3 of movies classified as profitable (1.24 ratio)
- A three class problem (profitable, not profitable, ambiguous)
 - Top 1/4 of movies classified as profitable, bottom 1/4 classified as not profitable

4. Data

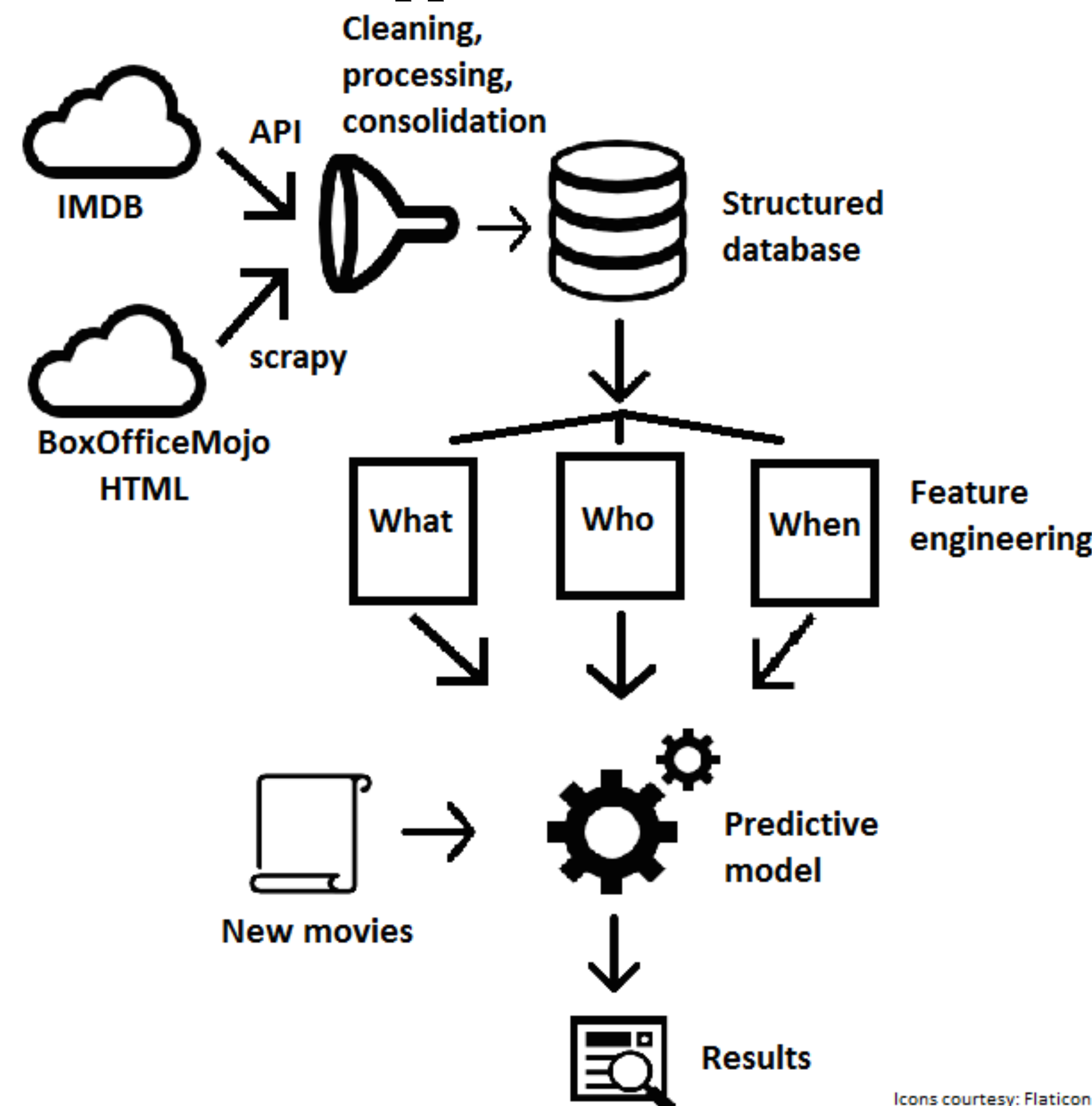
• Acquisition

- Scraped from BoxOfficeMojo.com and IMDb

• Period

- Movies for which we have profit data from 2000 - 2010

5. Decision Support



6. Results

a) Two class:

Random Forest (200 trees)			
	AUC	Accuracy	F-Measure
Both Class	0.711	0.713	0.822
Positive Class	0.711	0.176	0.27
Negative Class	0.711	0.946	0.655

b) Three class:

Random Forest (200 trees)			
	AUC	Accuracy	F-Measure
All Classes	0.748	0.608	0.572
Positive Class	0.803	0.48	0.52
Negative Class	0.679	0.219	0.314
Neutral Class	0.756	0.867	0.728

7. Attribute Information Gain

Rank	Attribute	Category
1	Avg Director Gross	Who
2	Total Director Gross	Who
3	Aggregate Actor Gross (total)	Who
4	Aggregate Actor Gross (avg)	Who
5	Aggregate Team Tenure	Who
6	Average Actor Gross (avg)	Who
7	Cumulative Actor Gross (top)	Who
8	Average Actor Gross (total)	Who
9	Cast Novelty	Who/What
10	Average Team Tenure	Who
11	Average Actor Gross (top)	Who
12	Avg. Weighted Genre Expertise	Who/What
13	Well-Rounded' Actor Appear.	Who
14	Average Genre Expertise	Who/What
15	Change in Clustering Coeff.	Who
16	Actor Cosine Similarity (max)	Who
17	Aggregate Betweenness Cent.	Who
18	Actor Cosine Similarity (avg)	Who
19	Average Degree (actor)	Who
20	"R" Rating	What
21	Annual Profit. % By Genre	When
22	Average Betweenness Cent.	Who
23	"Family"-oriented Plot	What
24	"Foreign" Genre	What
25	Annual Weighted Prof by Genre	When