What do Facebook Users Know About Movies?

Do they know things??

- Goal: predict critical reception using Facebook Likes
  - y = Metacritic score

Dataset

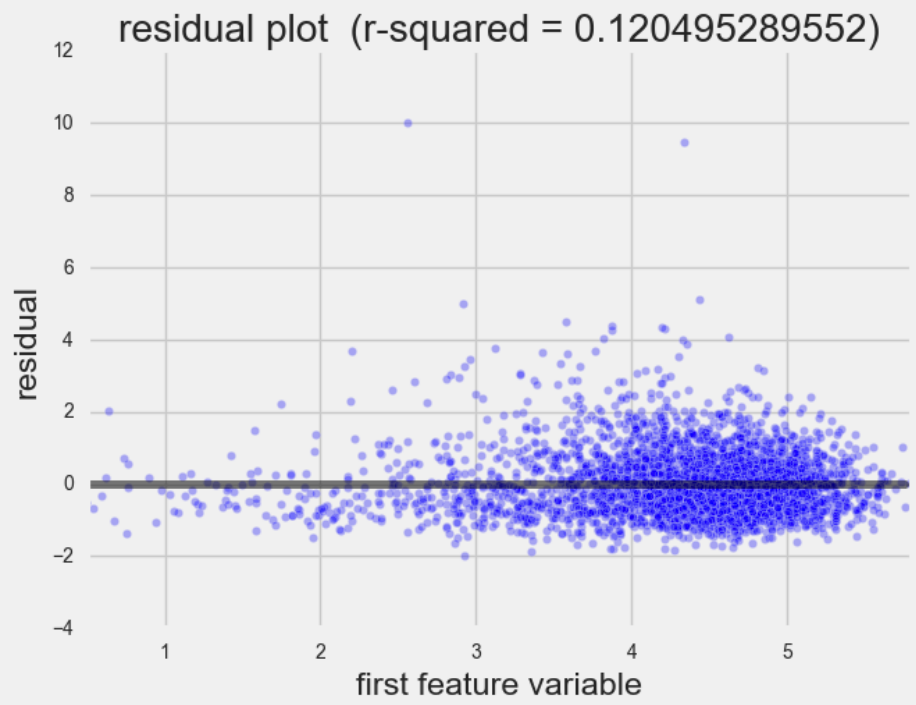
- 3200 movies titles from imdb.com
  - Metacritic score
  - Year of release
  - # of Facebook likes of the top 3 billed actors
  - # of Facebook likes of the director
  - # of Facebook likes of the writer
  - # of Facebook likes of the movie itself
- Preprocessing: Log[features] (except release year)

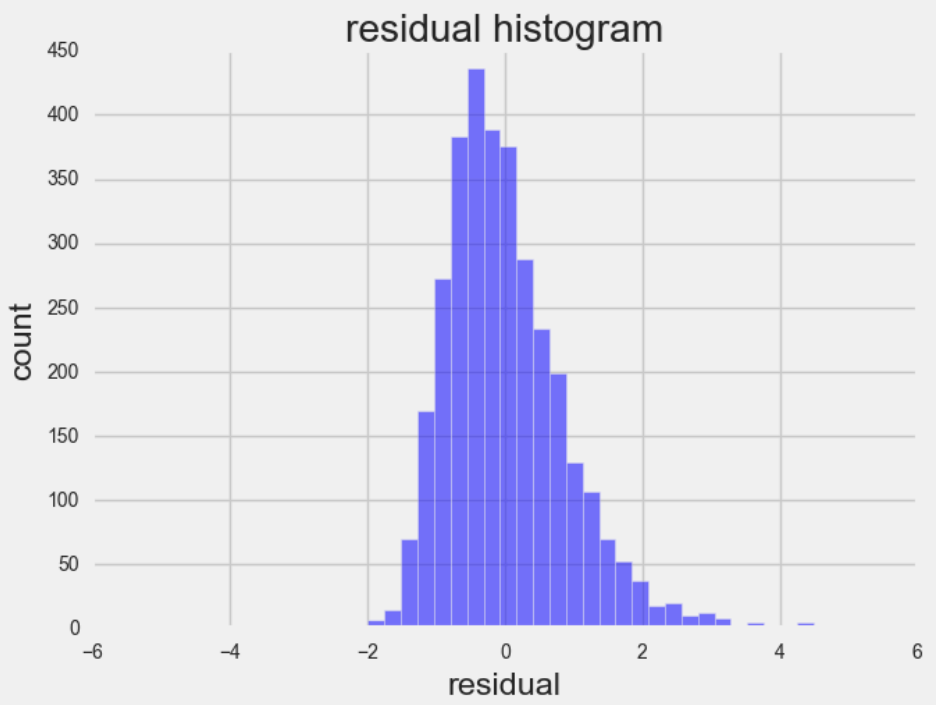
Naïve Model

- $y = b_0 + b_a X_{a\_likes} + b_w X_{w\_likes} + b_d X_{d\_likes} + b_m X_{m\_likes} + \varepsilon$
- Out-of-sample performance scored on 30% of the dataset set-aside initially
- Ridge regression, cross-validated on the training set (10 splits)

R-squared: FB1 – 0.120

Naïve Model





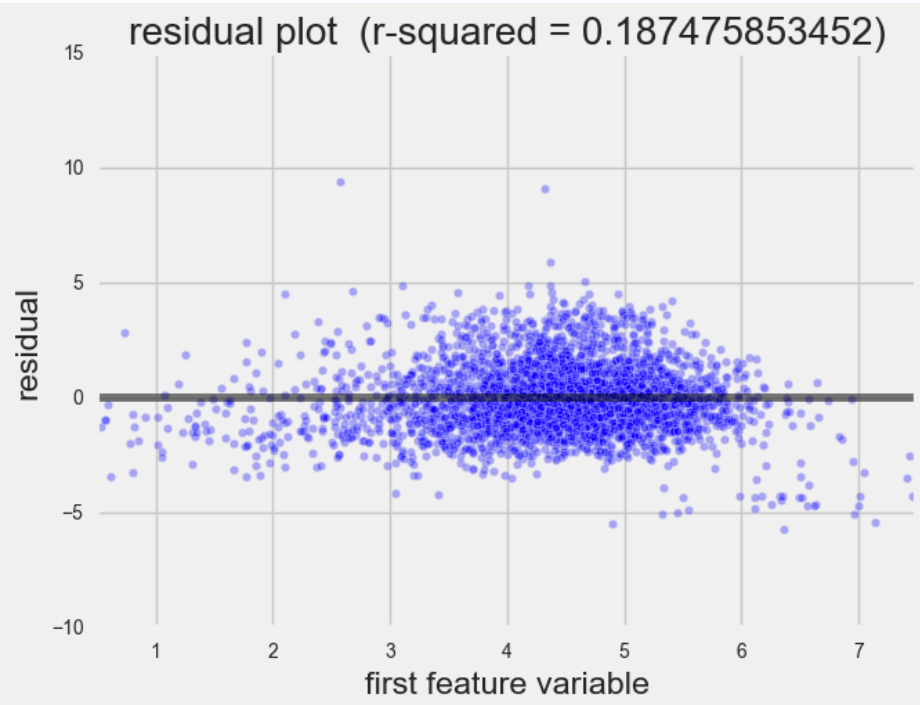
R-squared: FB1 – 0.120

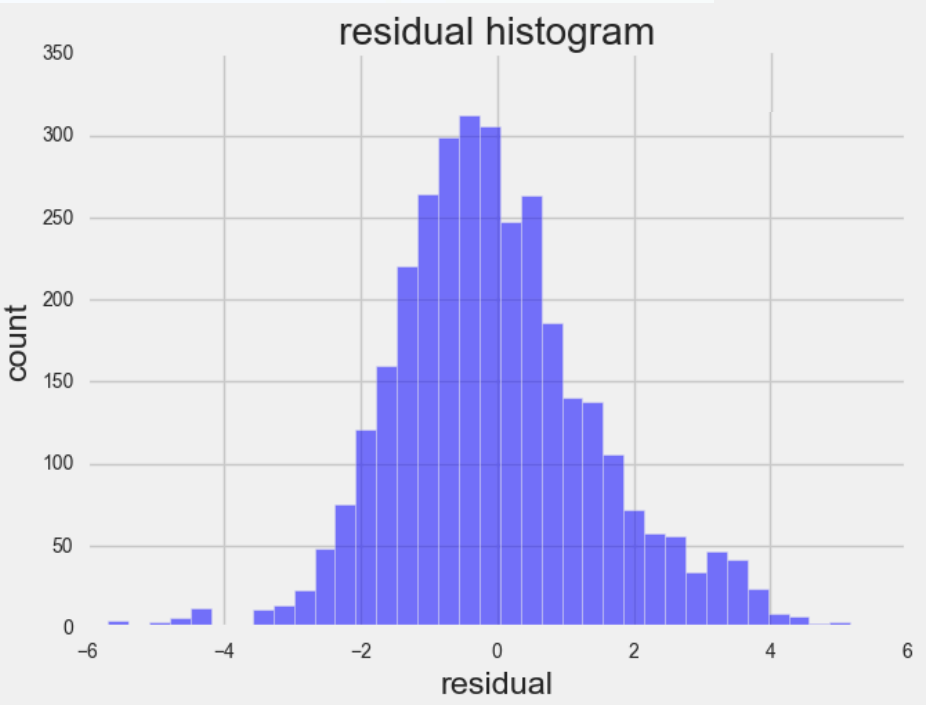
Group-Standardized Model

- Standardized data columns with **movies released on the same year**
- $y = b_0 + b_a X_{a\_likes} + b_w X_{w\_likes} + b_d X_{d\_likes} + b_m X_{m\_likes} + \varepsilon$
- Out-of-sample performance scored on 30% of the dataset set-aside initially
- Ridge regression, cross-validated on the training set (10 splits)

R-squared: FB1 – 0.120 | FB2 – 0.187

Group-Standardized Model

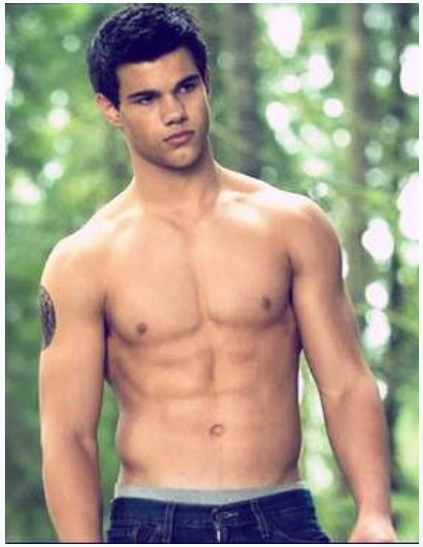




R-squared: FB1 – 0.120 | FB2 – 0.187

Conclusions so far…

- Insignificant features:
  - Writers ‘Likes’
    - Nobody cares about the writers
  - Actors ‘Likes’
    - People might not be  ing their acting abilities (e.g. they might be voting for abs)
- So what do Facebook users know about the movies?? Do they know things?
  - Not a whole lot…



R-squared: FB1 – 0.120 | FB2 – 0.187

Back to Scraping...

- Adjusted thesis:
  - How well can we **actually** predict the preference of movie critics?
- Hypothesis:
  - Critically-acclaimed directors, actors, writers make critically-acclaimed movies
- Additional features scraped:
  - Awards + Nominations of
    - Top 3 billed actors
    - Directors
    - Writers

- Same processing as before:
  - Log[features]
  - Grouped by year and scaled

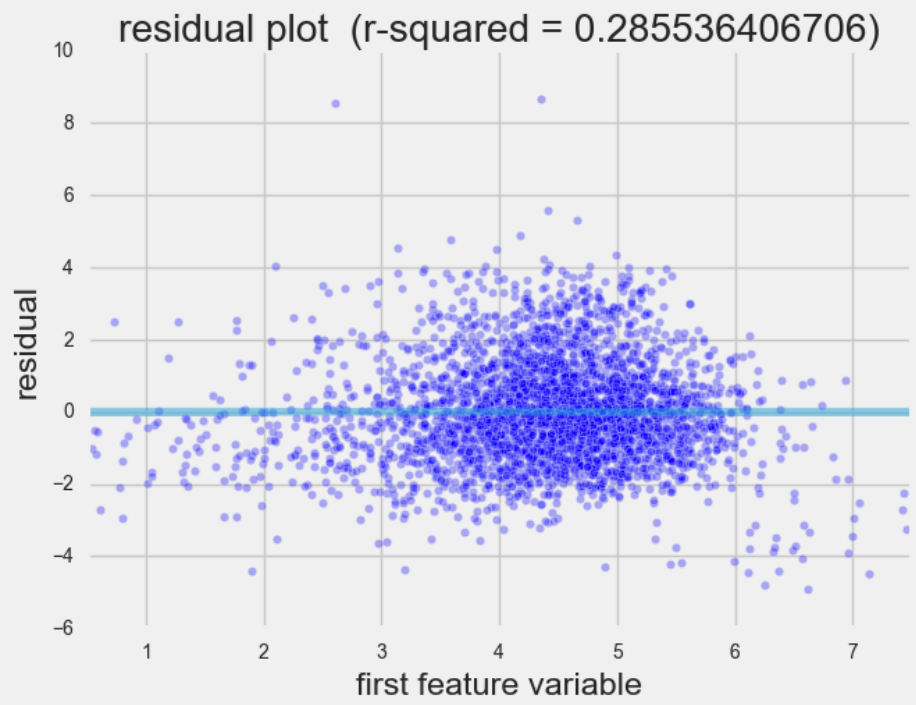
R-squared: FB1 – 0.120 | FB2 – 0.187

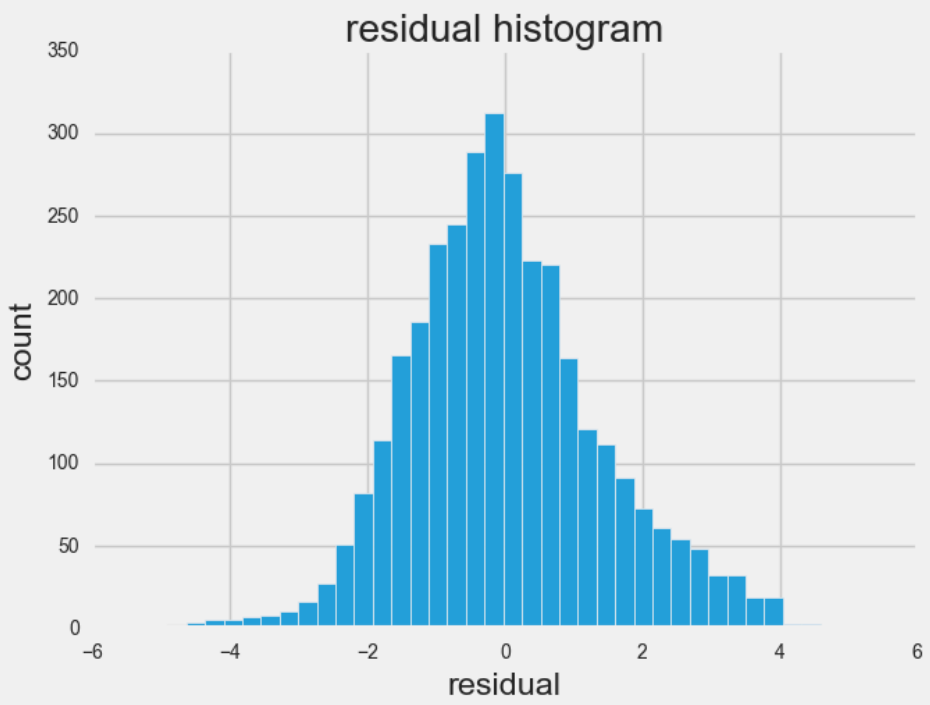
New Model (with awards)

- $y = b_0 + b_{aL}X_{a\_likes} + b_{wL}X_{w\_likes} + b_{dL}X_{d\_likes} + b_{mL}X_{m\_likes}$
  $\mathbf{+ b_{aA}X_{a\_awards} + b_{wA}X_{w\_awards} + b_{dA}X_{d\_awards}} + \varepsilon$
- Out-of-sample performance scored on 30% of the dataset set-aside initially
- Ridge regression, cross-validated on the training set (10 splits)

R-squared: FB1 – 0.120 | FB2 – 0.187 | AW1 – 0.286

New Model (with awards)

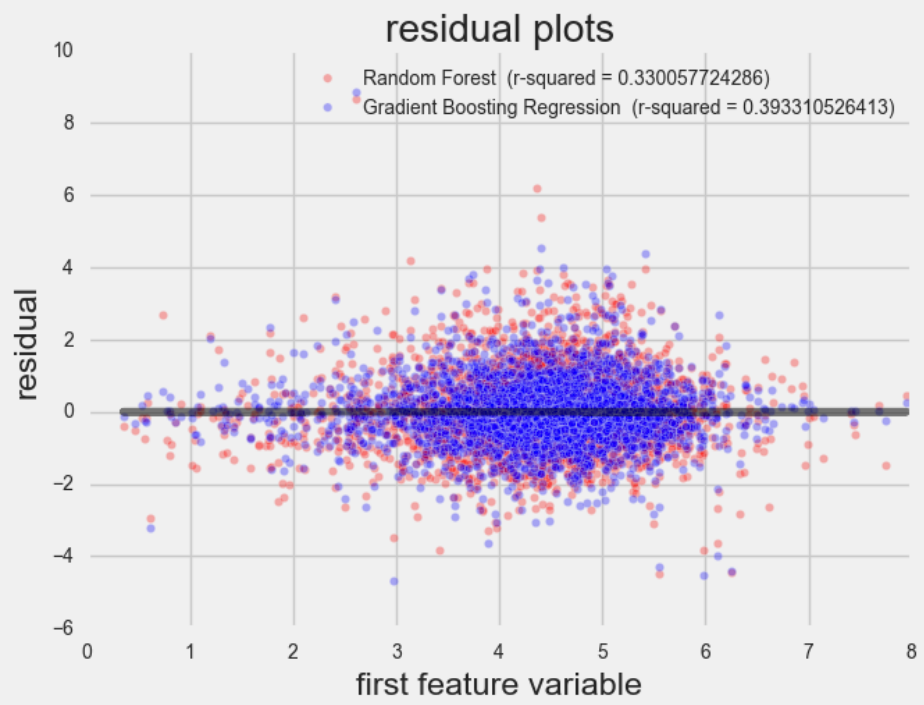


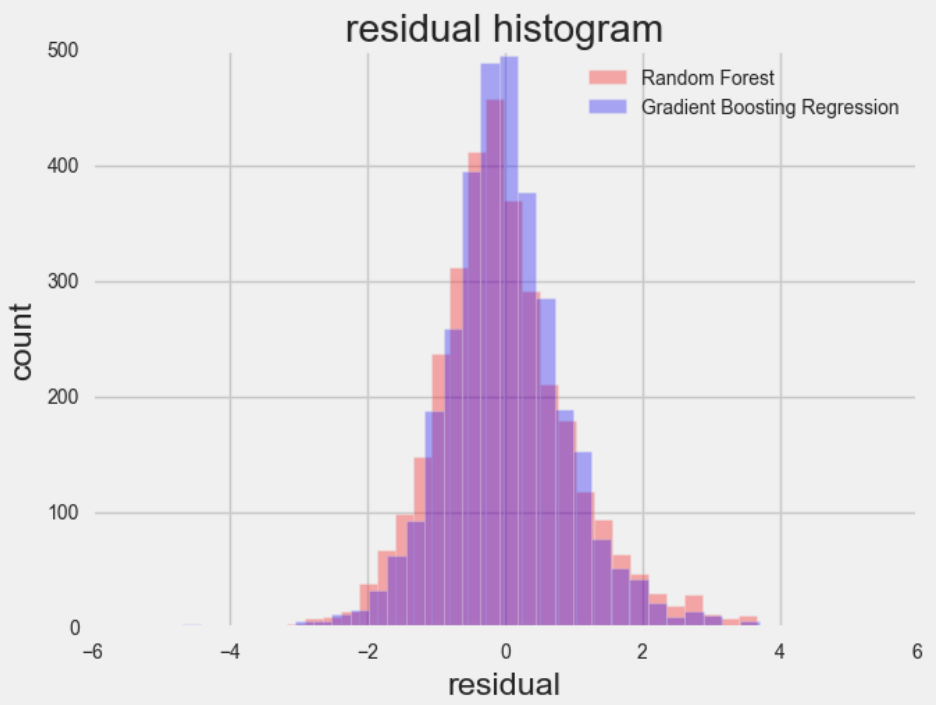


R-squared: FB1 – 0.120 | FB2 – 0.187 | AW1 – 0.286 | AW2,3 – 0.330, 0.393

Ensemble Methods (with awards)

- Random Forest (r-squared ~ 0.330)
- Gradient Boost (r-squared ~ 0.393)





Final Conclusions

- Facebook knows nothing
- Critically-acclaimed directors, actors, writers **don't always mean** critically-acclaimed movies
  - See Ocean's Twelve...
- Critics might just be unpredictable rogues...
  - It would be nice if Metacritic published variance of the reviews for each movie too!
- Ensemble methods = voodoo magic