

The Automated Script Reviewer

Exploring the relationship
between scripts and ratings

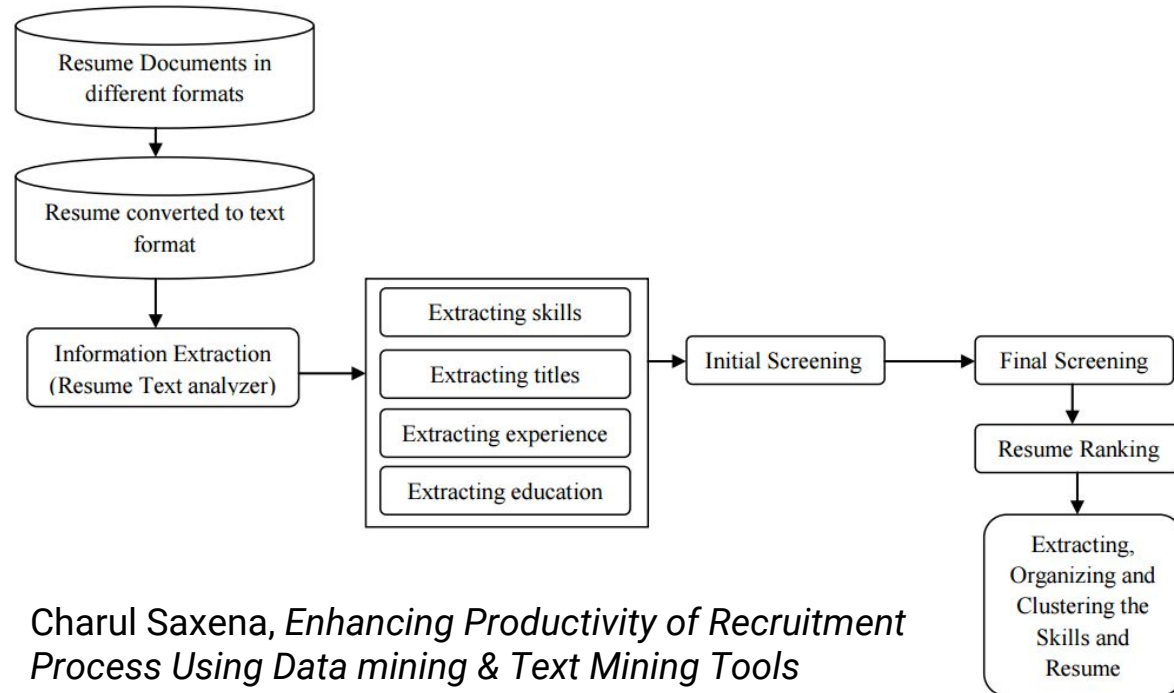


F.R.I.E.N.D.S



Katherine Schinkel & Marcus Rosti

The Resume Reviewer



Can this method work for television scripts?

Pivotal Data Labs

*Using Data Science
to Predict TV
Viewer Behavior
and Formulate a Hit
TV Show*

Nielsen Ratings

Show Transcripts

Manually Collected
Metadata

Parsing
Tokenization
Lemmatization
TF-IDF
Corpus Reduction
LDA Topic Modeling

Linear
Regression

“the most relevant variables were derived from the transcript data”

<https://blog.pivotal.io/data-science-pivotal/case-studies/using-data-science-to-predict-tv-viewer-behavior-and-formulate-a-hit-tv-show>

Data Collection

TV Scripts



175 episodes

F.R.I.E.N.D.S

207 episodes

383 Documents



Ratings

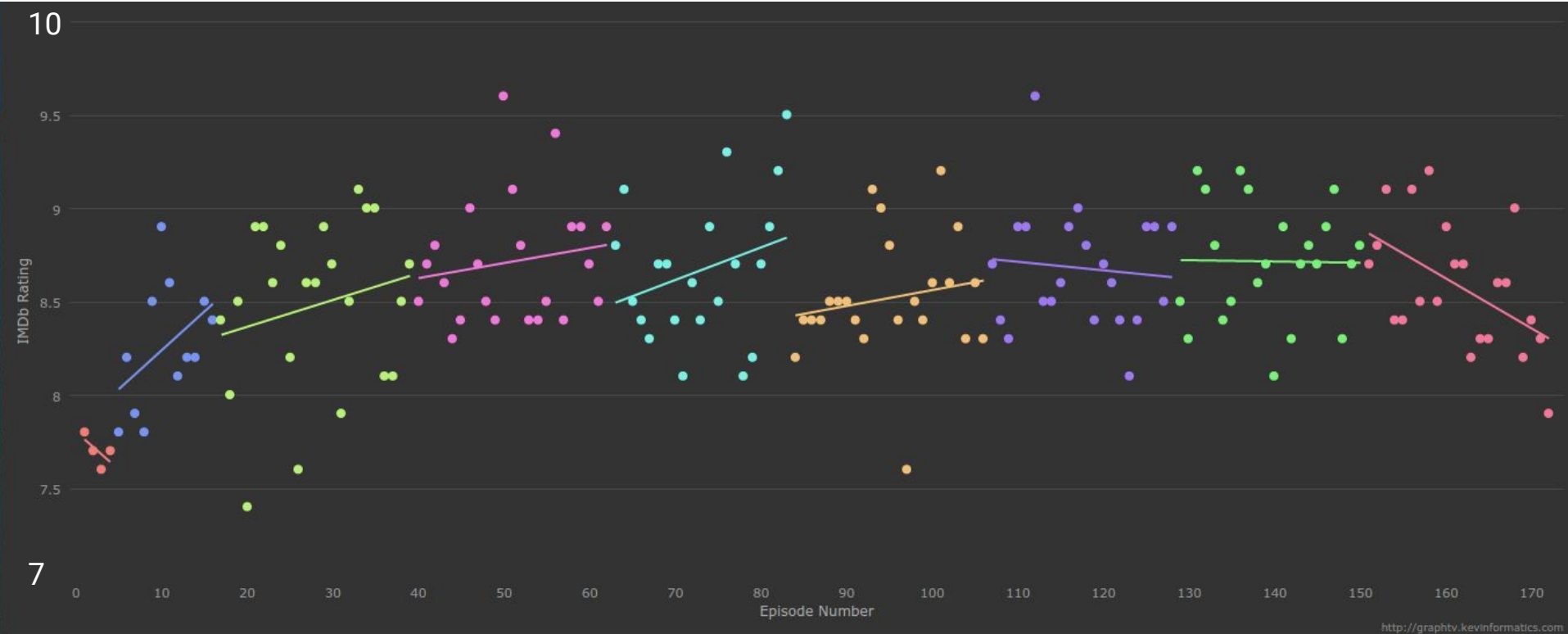


10 pt. Rating Scale

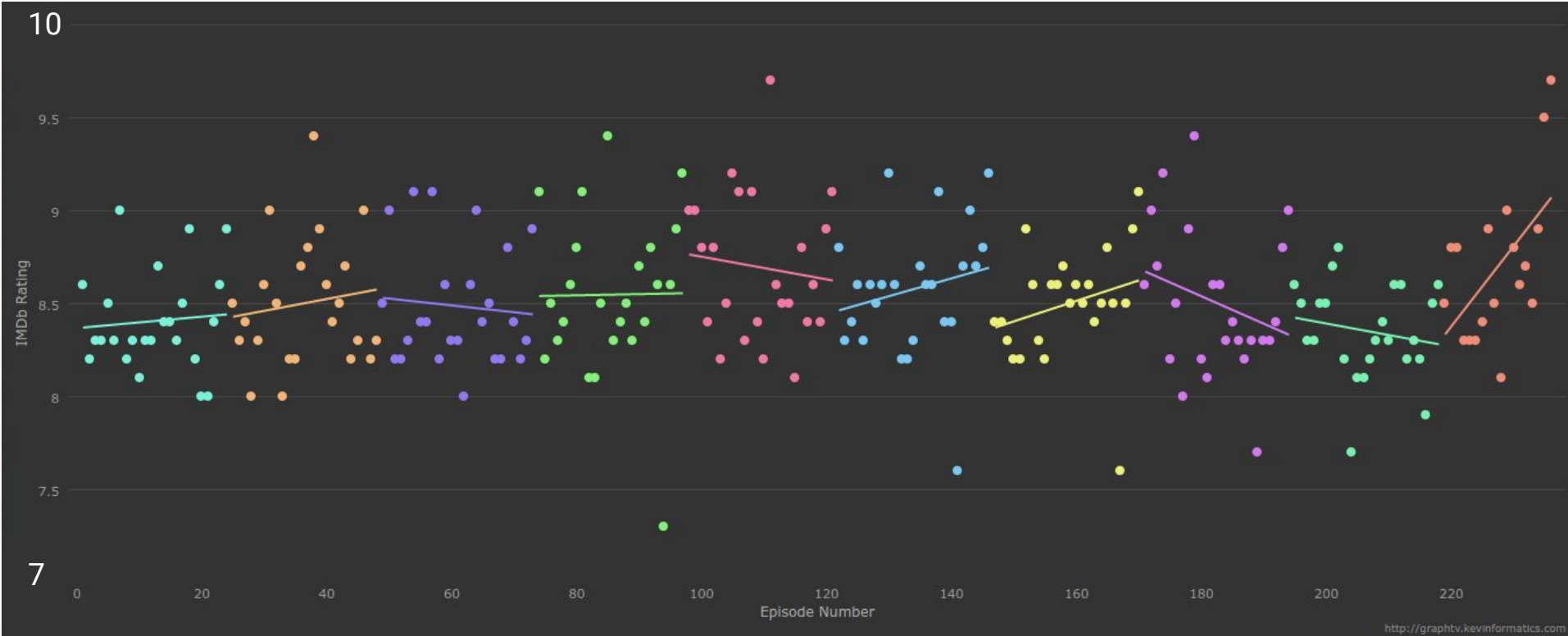


Nielsen Ratings

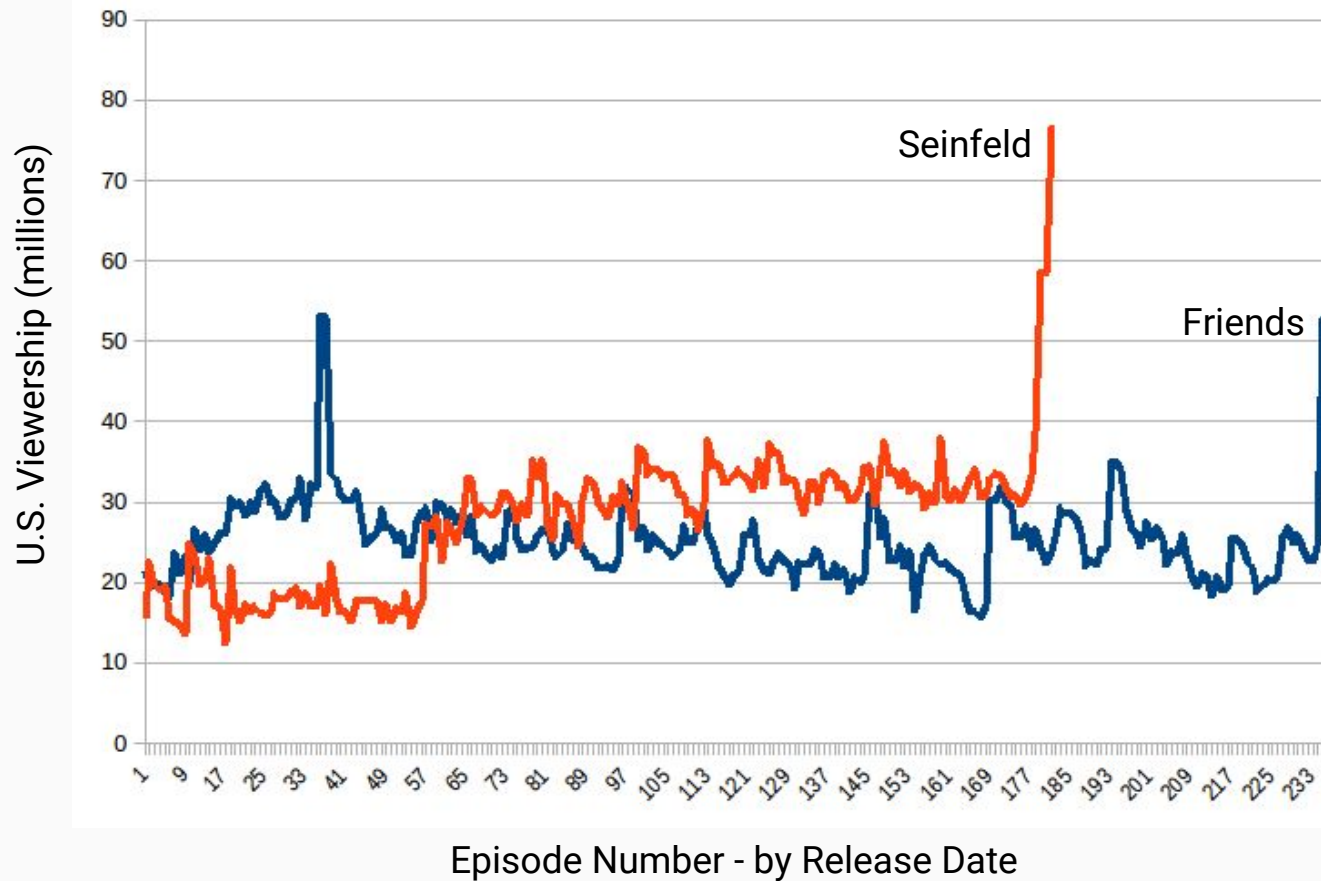
Seinfeld IMDb Ratings



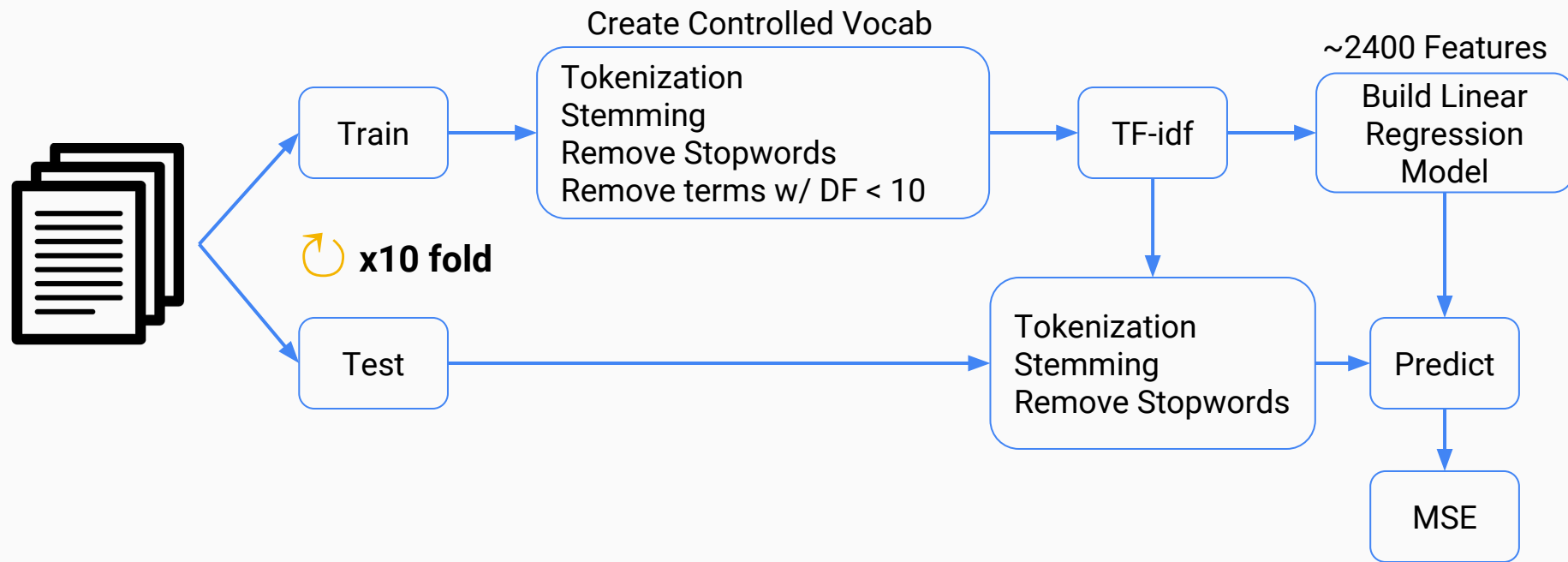
Friends IMDb Ratings



Nielsen Ratings



Base Modeling Approach



Base Modeling Evaluation



10 pt. Rating Scale



Mean MSE = 0.134
Model MSE = 0.175
t = -1.713

Paired t-test

$$\frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\sigma_1^2/n + \sigma_2^2/n}}$$

H_0 : Model MSE = Mean MSE
 H_A : Model MSE < Mean MSE

We want t > 1.812

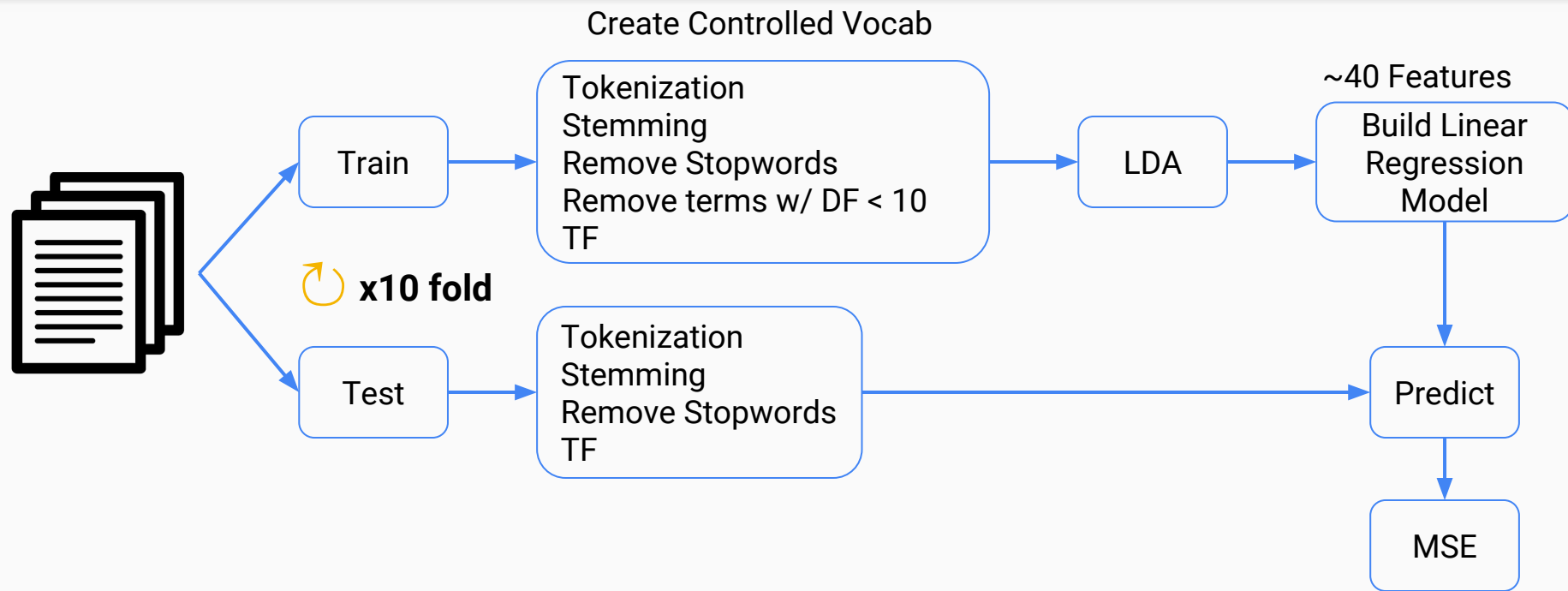


Nielsen Ratings



Mean MSE = 50.018
Model MSE = 70.269
t = -1.226

LDA Modeling Approach



LDA Modeling Evaluation



10 pt. Rating Scale



Mean MSE = 0.133
Model MSE = 0.132
t = 0.062

Paired t-test

$$\frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{\sigma_1^2/n + \sigma_2^2/n}}$$

H_0 : Model MSE = Mean MSE
 H_A : Model MSE < Mean MSE

We want t > 1.812



Nielsen Ratings



Mean MSE = 45.858
Model MSE = 44.872
t = 0.096

Possible Improvements

Additional Feature Engineering (scene count, characters per scene count, characters per episode count)

Additional Modeling Techniques (Regression Trees, Lasso, ElasticNet Regression)

Consider multicollinearity of regression features

Use a classification approach with binning

Incorporate bigrams and/or trigrams

Questions?