

# ONLINE NEWS POPULARITY

*Evaluation and Prediction*

---

Zhiyi Guo, Wenyu Ma, Sili Wang, Evelyn Zheng, Xintong Zhou

# OBJECTIVES

---

- Find the best model to predict the popularity of online news
- Identify the set of features that have major influence on popularity
- Achieve better website design to attract more readers

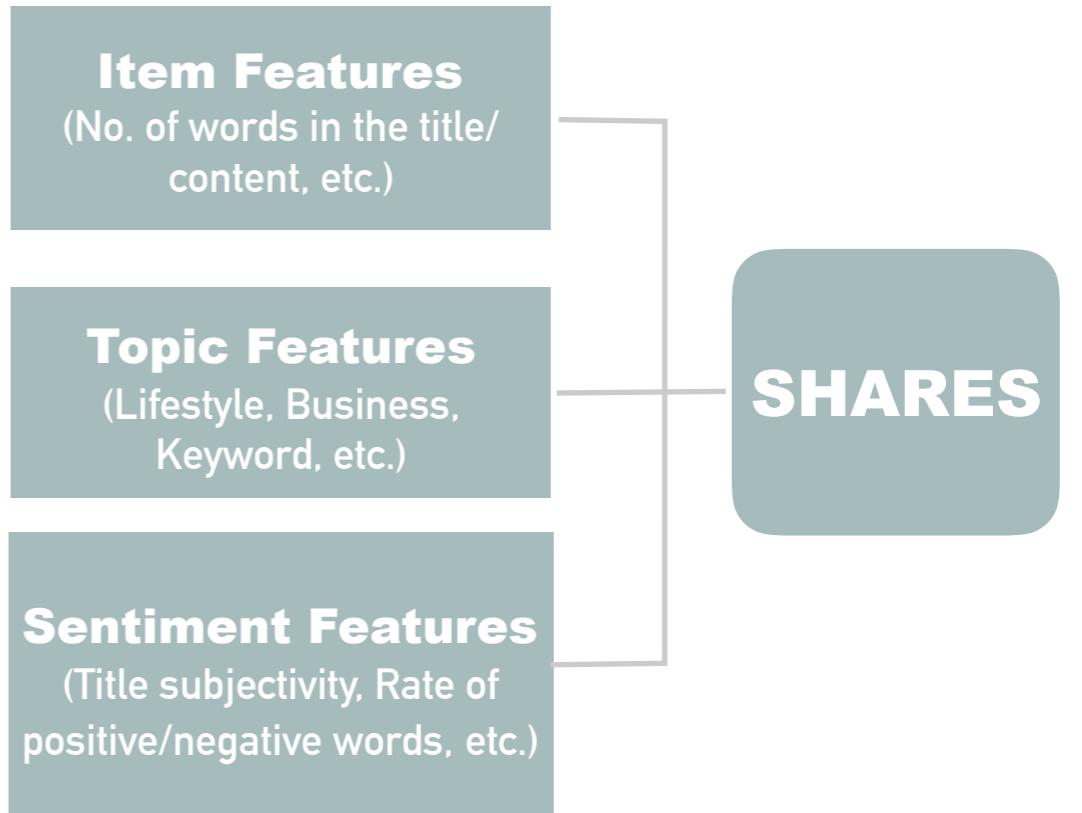
# SOURCE AND BACKGROUND

---

- Source: UCI Machine Learning Repository
- Variables: 58 predictors + 1 response (shares)
- Sample size: 37773 articles from mashable.com in 2013-14

# DATA

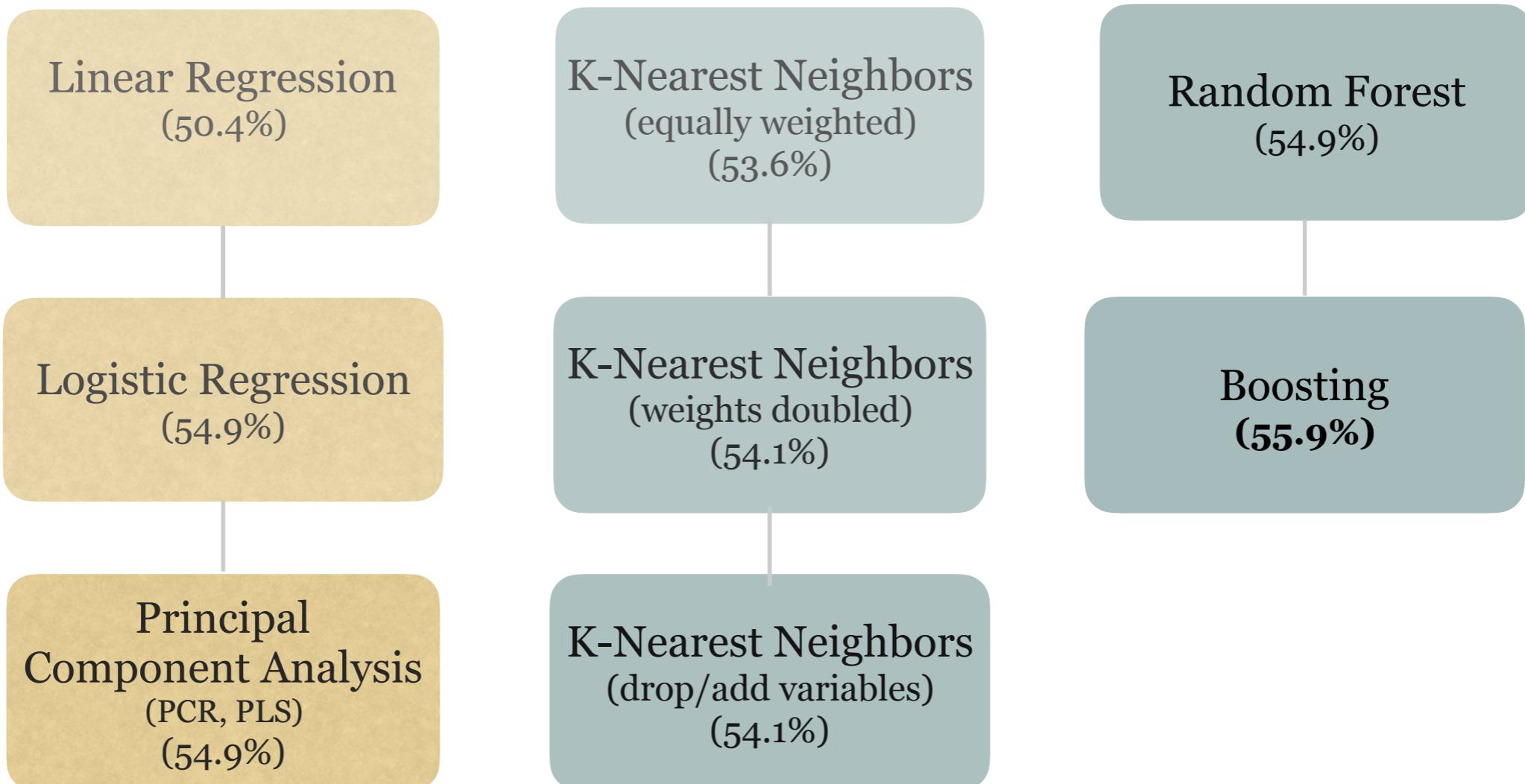
---



# METHODOLOGY

---

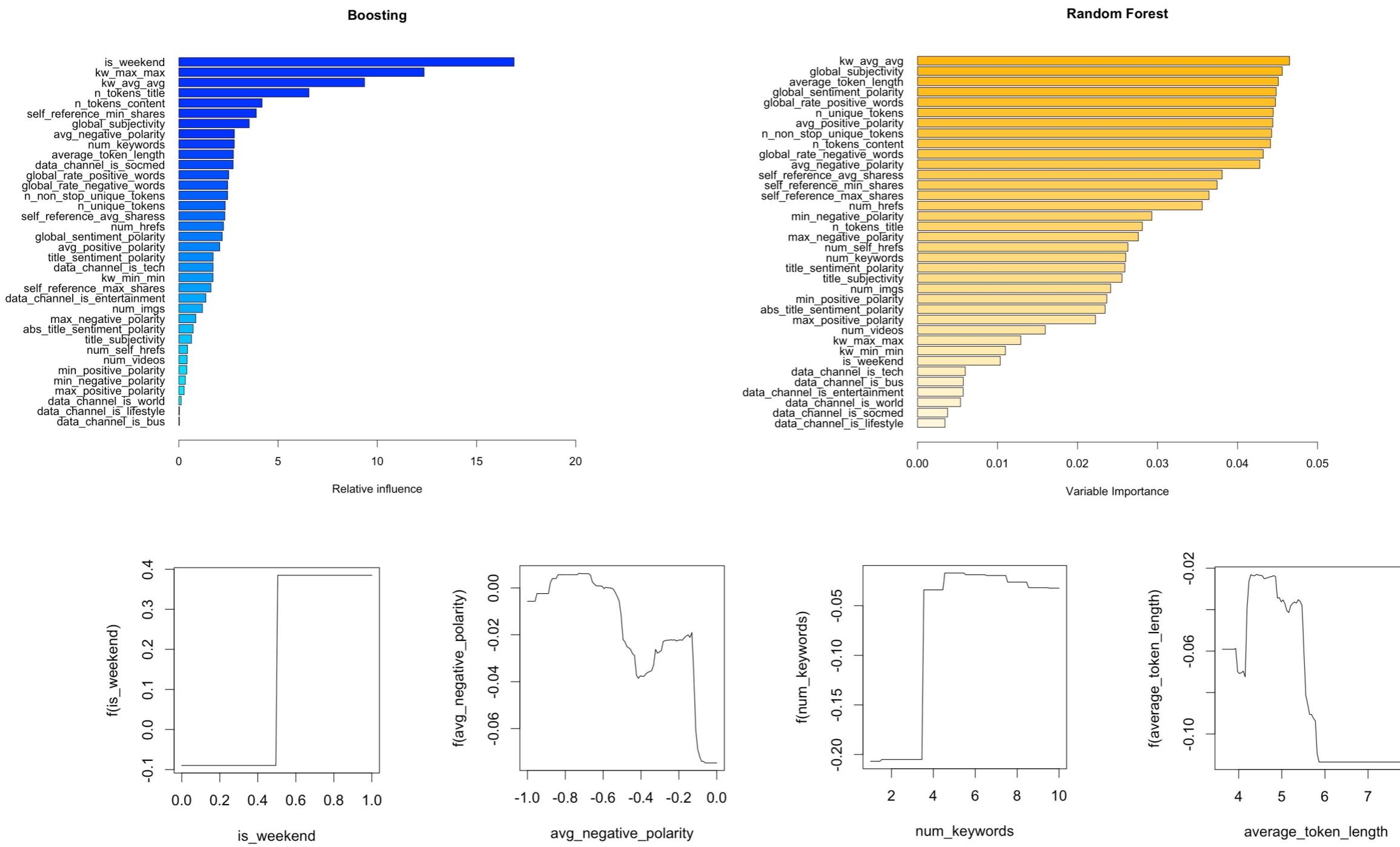
## ► Model selection and advancement



# RESULTS

---

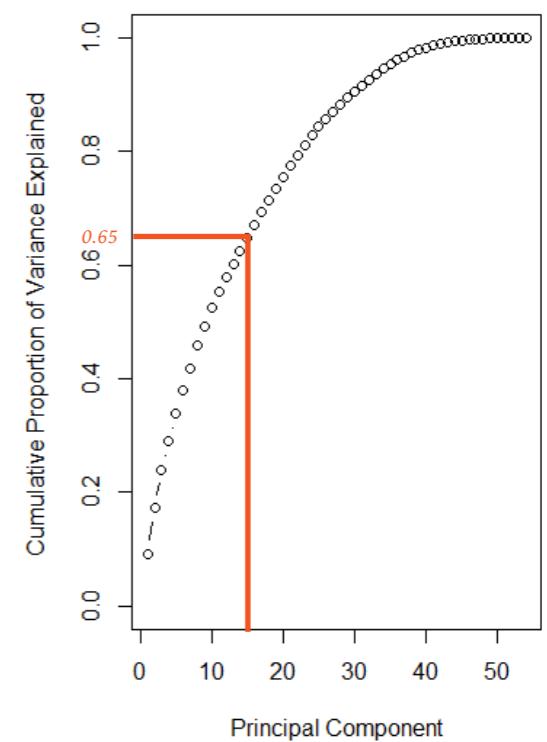
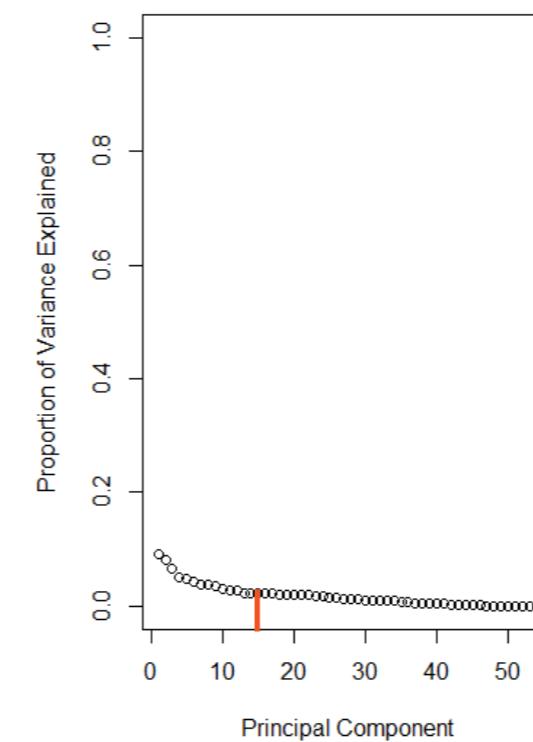
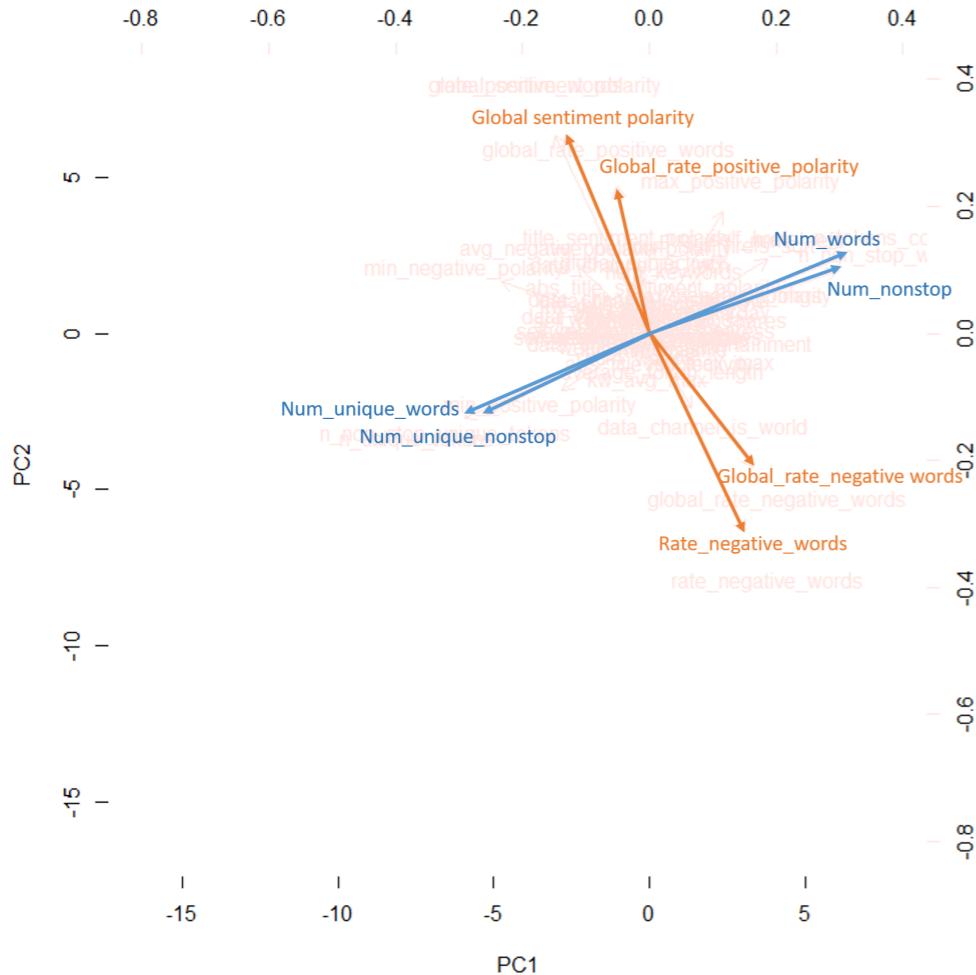
- Boosting achieved a highest true positive rate of 72.7%; Random Forest achieved 63.5% true positive rate



# RESULTS

---

- PCA for dimension reduction
- PC1: Content-related (negative)
- PC2: Emotional-related (positive)
- Unsupervised learning : PCR, PCR logistic
- Supervised learning : PSL, PSL logistic, PC for KNN



# CONCLUSIONS

---

## Achievements

- Dimension reduction
  - A tailored attributes selection
  - Weekend publication indicator
    - Evergreen vs. time-sensitive articles
  - Tone is important
- 

A dog dressed as a bear on the treadmill will get you through this week

211.6k  
SHARES



# RECOMMENDATIONS

---

## Areas for Improvement

- Why the prediction results are not ideal
  - Audience population not representative
  - Lack of crucial attributes
  - treadmill example
- A equally-weighted model may not be good start for KNN
- Possible attributes to look at
  - Video details
  - Title word choices
  - Trending topic indicator