



How can we extract information from language?



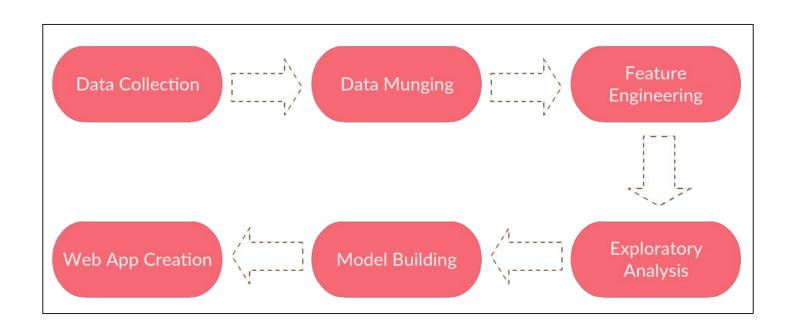
How can we extract information from language?

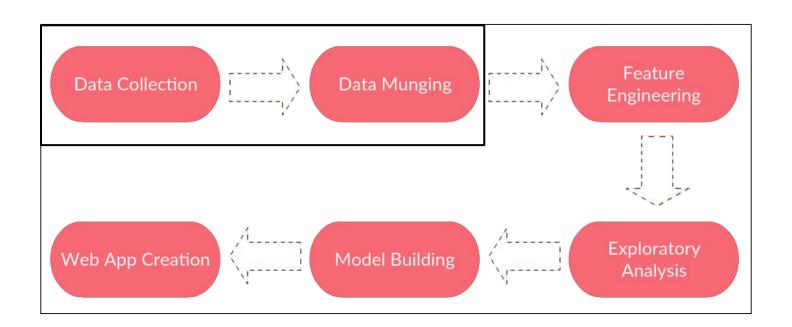
body language () intonation () expressions



Is there a "body language" of written text?

Can it be used to predict information about the writer?











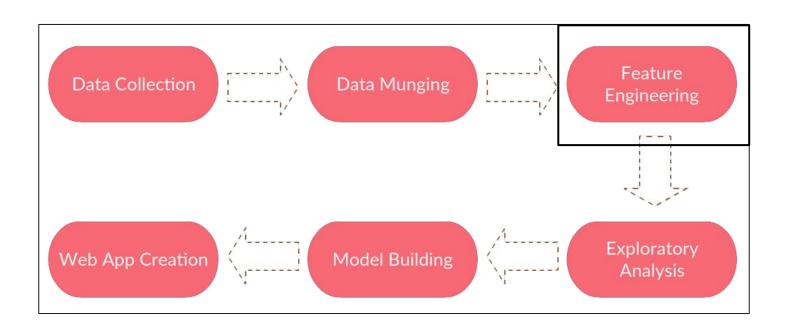
BuzzFeed



The New York Times

Slate

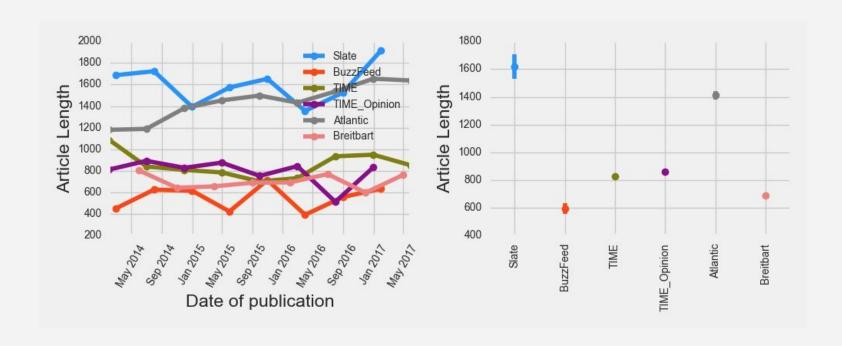
Data collection





Feature Engineering

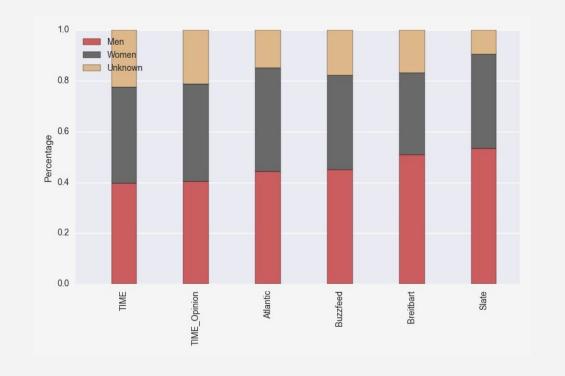
- Mean token length
- Frequency of punctuation
- Subjectivity/Polarity
- etc. ...



News sites exhibit drastic differences in stylistic features.

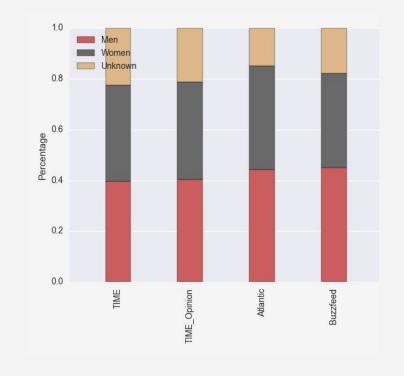


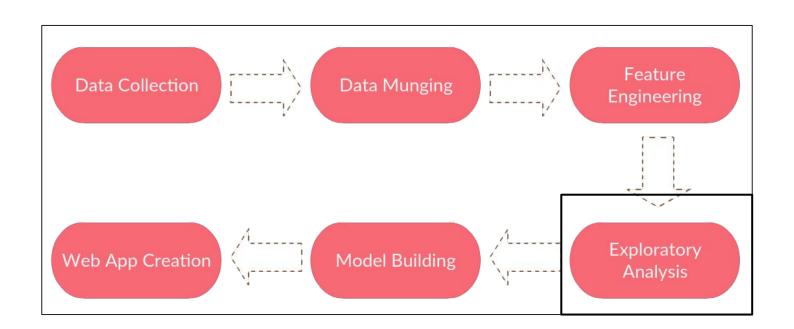
dedia Sites Gender Ratio

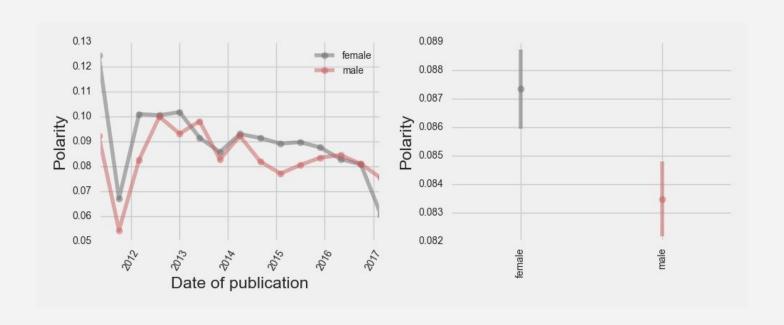




Media Sites Gender Ratio







Females and males exhibit strong differences for many stylistic features.



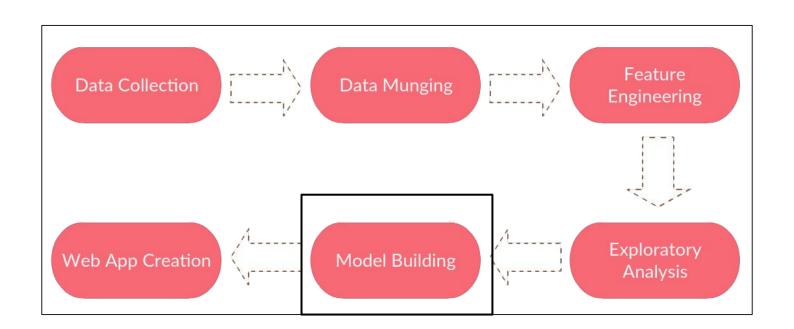
TOPIC MODELING? NOT GENDER MODELING?

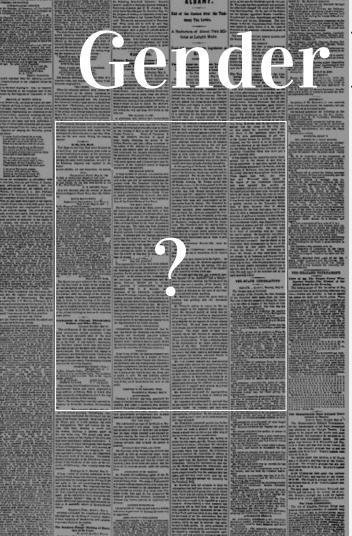


FEMALE: 'one', 'people', 'new', 'like', 'time', 'many', 'could', 'years', 'women', 'even', 'u.s', 'president', 'state', 'first', 'house'

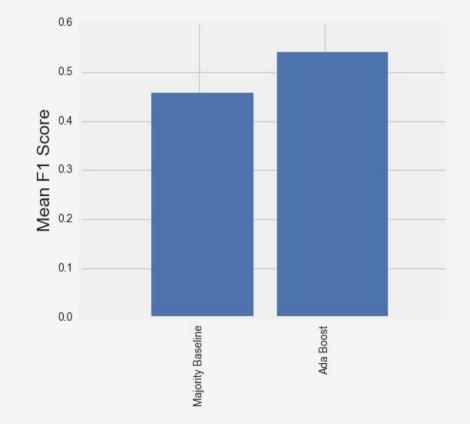
MALE: 'one', 'people', 'new', 'like', 'time', 'u.s', 'could', 'president', 'even', 'years', 'many', 'government', 'state', 'states', 'first'

Most used words



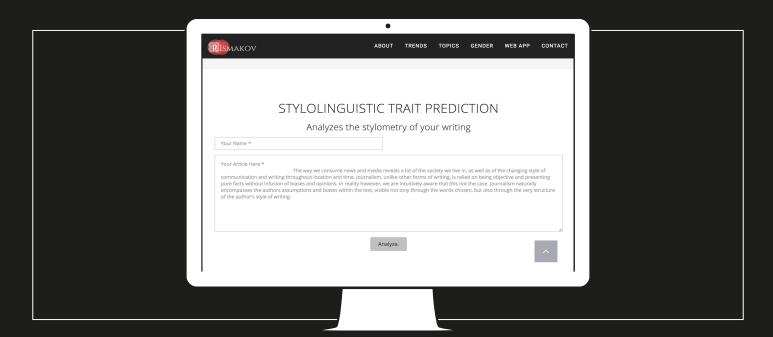


Predictions

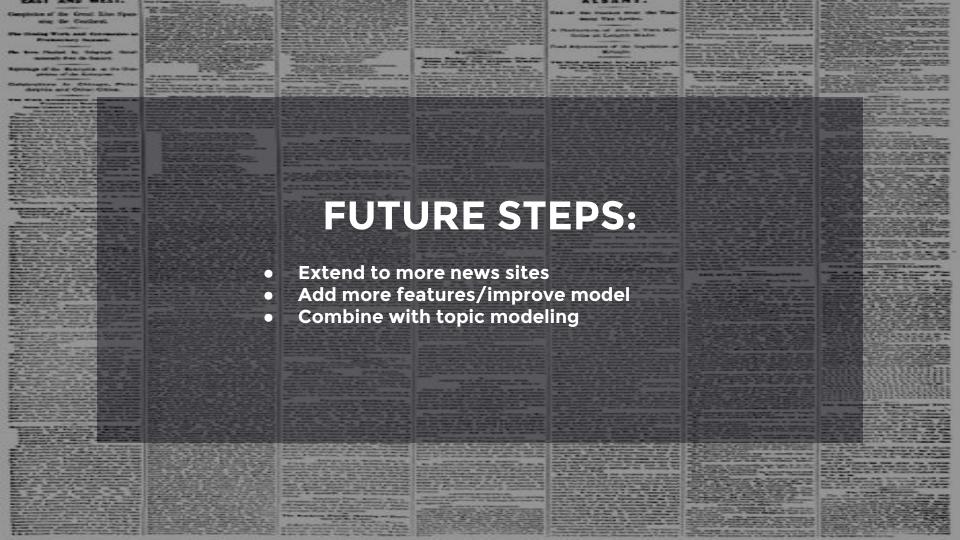


WEB APP PROJECT

Put in your text or article and will output information and predictions.



rismakov.com/app





THANKS!

Any questions?

You can reach me at

rismakov@gmail.com

github.com/rismakov/Lexophilia

-Revekka Ismakov-