WEB API & CLASSIFICATION MODEL

JEREL NOVICK

JULY 11, 2019

OUTLINE

- DAD JOKES
- DATA COLLECTION
- MODELS
- CONCLUSIONS
- FURTHER WORK
- QUESTIONS

DAD JOKES

- CORNY
- GENERALLY UNFUNNY
- ONE LINER
- REMINISCENT OF REMARKS MADE BY MIDDLE AGED OR ELDERLY FATHERS
- INVOLVES A PUN

Sources: Wikipedia and Dictionary.com

DAD JOKES

DATA COLLECTION

- 17 SCRAPES
- JOKES
  - 2,509 Posts
- DAD JOKES
  - 1,648 Post

MODELS

- LOGISTIC REGRESSION
- GRID SEARCH
- NAÏVE BAYES

LOGISTIC REGRESSION MODEL

- 2,000 MAX FEATURES
- TRAIN SCORE 0.85338
- TEST SCORE 0.59807

	Predicted Negative	Predicted Positive
Actual Negative	434	161
Actual Positive	257	188

GRID SEARCH MODEL

- TRAIN SCORE 0.85338
- TEST SCORE 0.59807

	Predicted Negative	Predicted Positive
Actual Negative	431	164
Actual Positive	239	206

NAÏVE BAYES

- TRAIN SCORE 0.71671
- TEST SCORE 0.58461

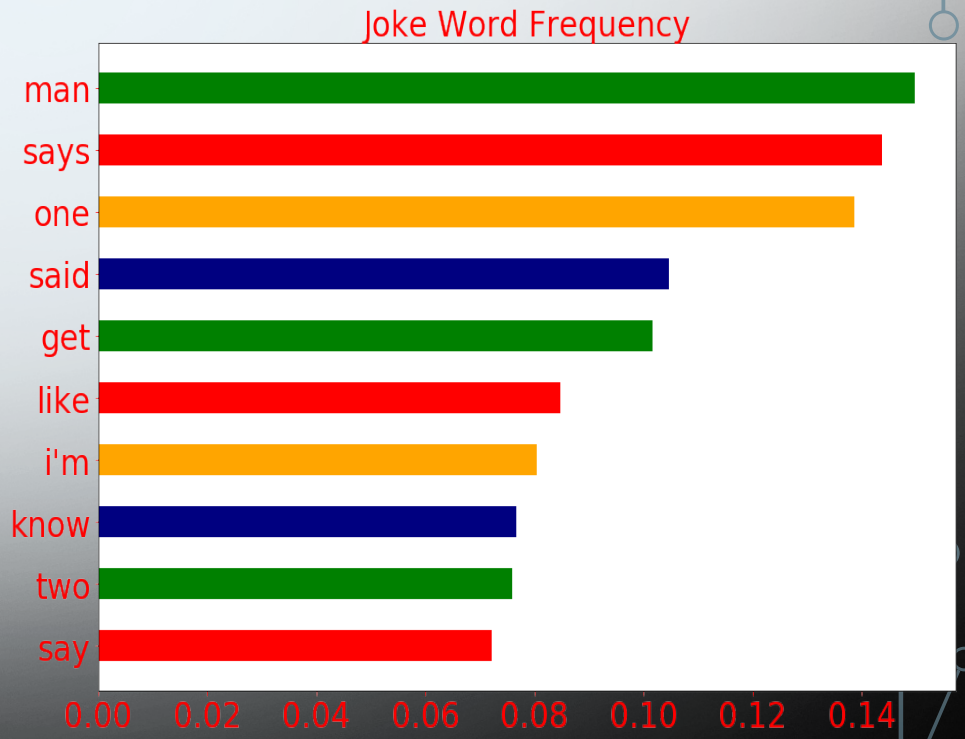
	Predicted Negative	Predicted Positive
Actual Negative	380	215
Actual Positive	217	228

CONCLUSION

CONCLUSIONS

- TRAINING MODELS OVERFIT
- GENERALLY BETTER AT PREDICTING THE NEGATIVE
- BIASED

FURTHER WORK



FURTHER WORK

	Absolute Coefficients
Whole	1.5740
Cow	1.5239
Sh*t	1.5099
Talk	1.4071
Middle	1.3700
Completely	1.2928
Largest	1.2807
Magician	1.2795
Running	1.2766
Happens	1.2714

FURTHER WORK

- MORE SELECTIVE FEATURES
- COMPARE AGAINST OTHER SUBREDDITS
- TONE
- PUNS

QUESTIONS