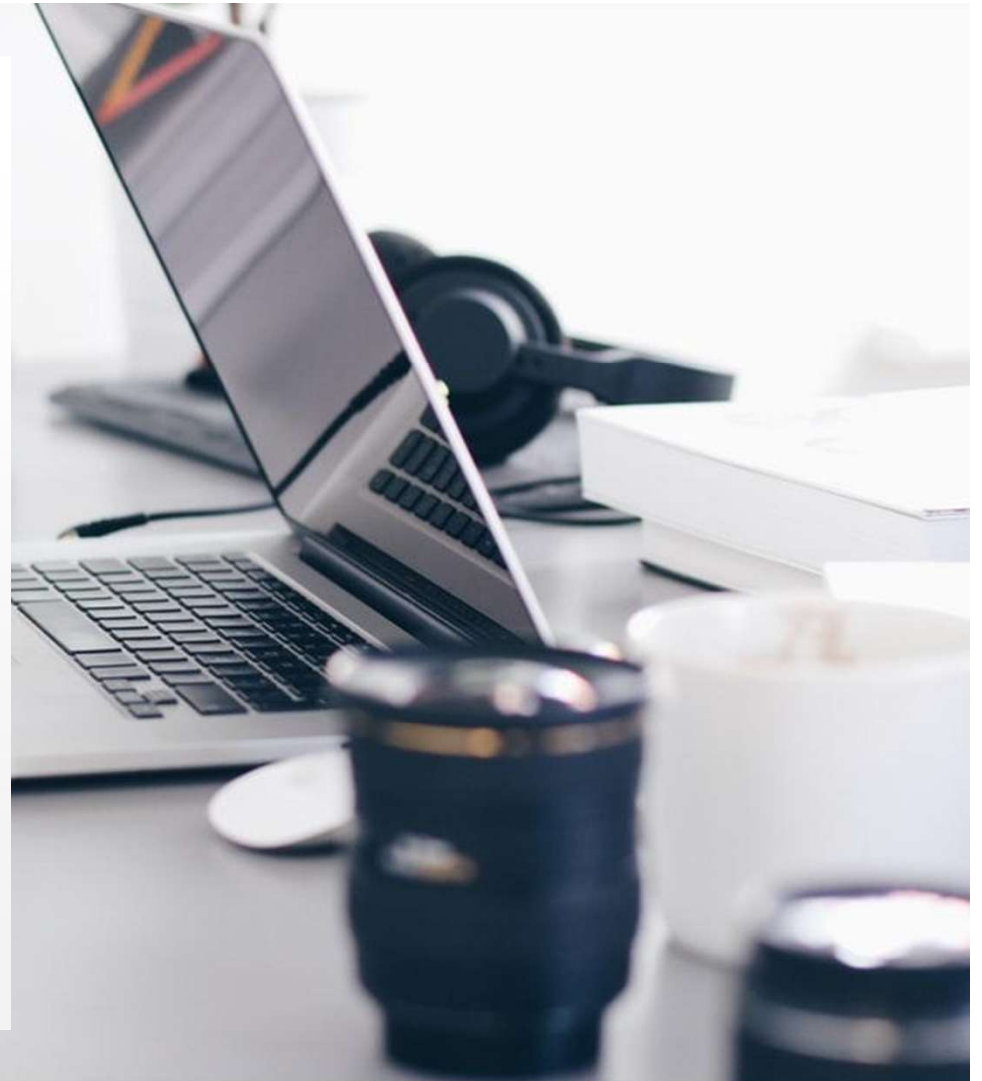


# **Classification Project – American Trends**

**Chris Byrnes  
May 14, 2021**

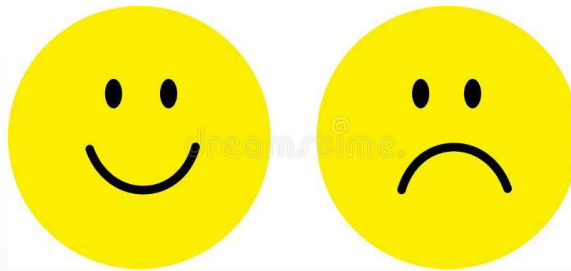


# Data

- Pew Research Center
  - Nonprofit and nonpartisan
  - Publish facts for dialogue and decision making
- American Trends Panel
  - Wave 67
  - April 29 – May 5, 2020

# Data

- Wave 67
  - Covid
  - Climate change
  - US Government responsibilities
- Data for 9,955 respondents out of 10,957 was used
  - Missing values for difference



# Project

Predict response to survey question: *Do you think a year from now, May 2021, will be better or worse for you than May 2020?*

# Baseline - Simple KNN with K = 1

## Confusion Matrix

		Predicted	
		Better	Worse
Actual	Better	1,548	189
	Worse	208	46

Accuracy	.801
Precision	.196
Recall	.181

## Issues / Objectives

- Imbalance
  - Total population – 13.6% w/ positive target
  - Test data – 12.8% w/ positive target
- Preferred metric – Recall
- Interpretable model

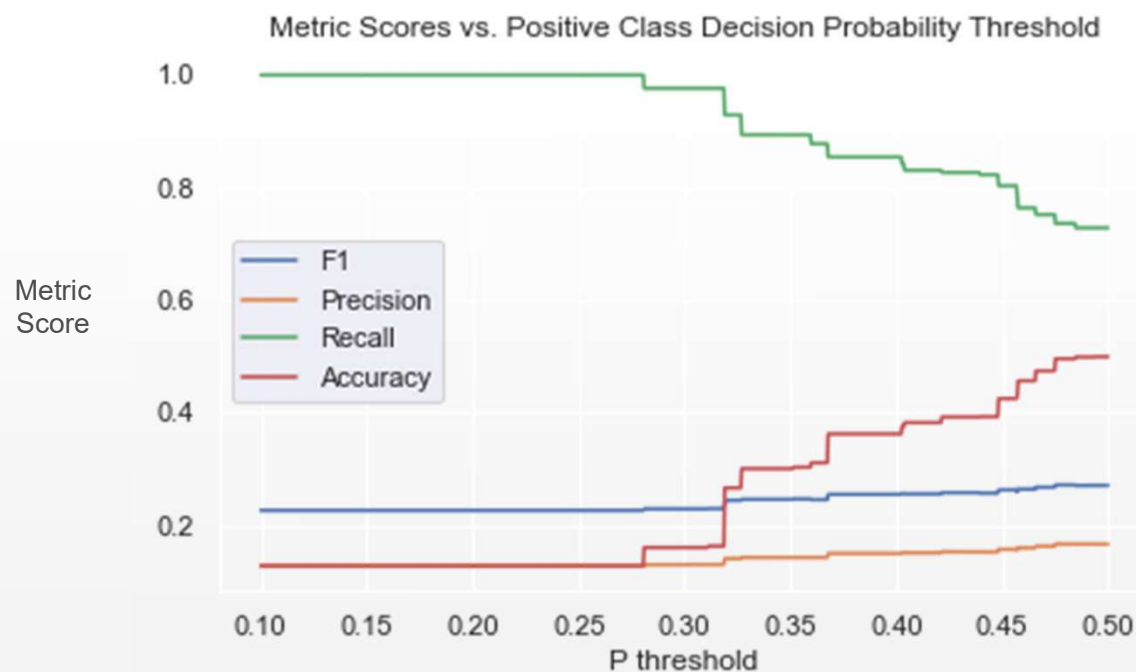
## Final Model – Logistic Reg., Class Weight = Balanced

### Confusion Matrix

		Predicted	
		Better	Worse
Actual	Better	809	928
	Worse	69	185

Accuracy	.499
Precision	.166
Recall	.728

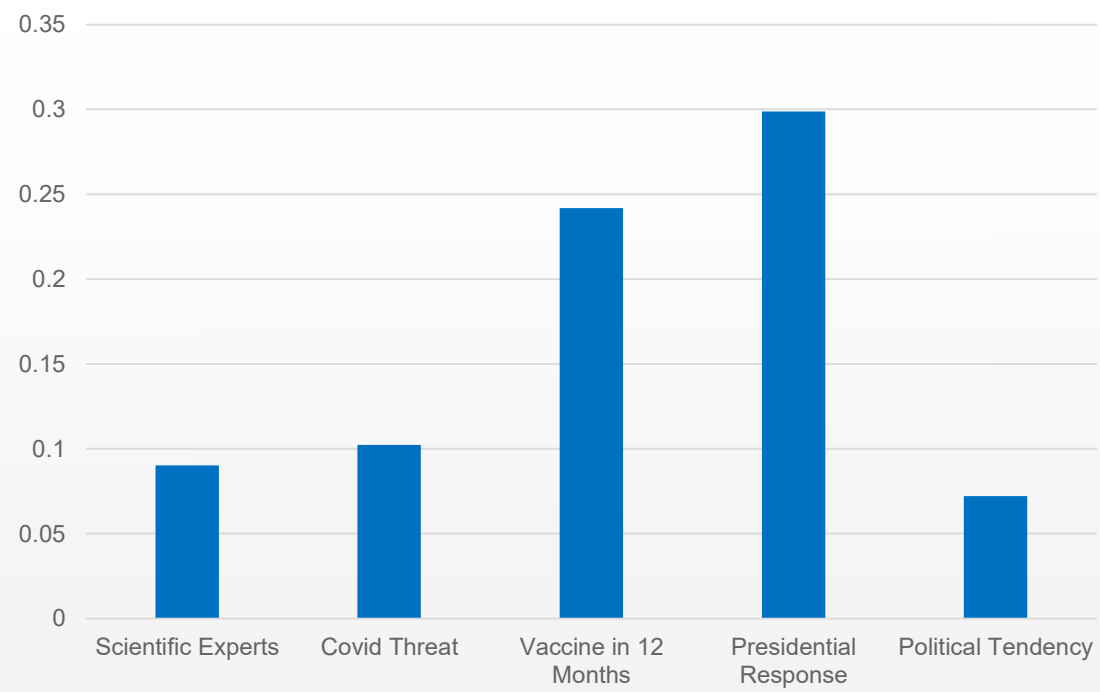
# Final Tweak – $P = .45$



Best F1 score is at .475 Threshold. Lowering P to .45 does not impact precision significantly and raises recall to .803



# Feature Importance



## Appendix – Survey questions used in final model

- In general, would you say scientific experts are usually \_\_\_\_\_ at making policy decisions about scientific issues than other people.
- How much of a threat is, if any, is the coronavirus outbreak for the health of the US population as a whole?
- How would you rate the job the following is doing responding to the coronavirus outbreak: the president?
- Thinking about the year ahead, that is in the next 12 months, do you think there will be a vaccine to prevent the coronavirus disease?
- Also used was political party tendency which was not part of the survey but provided with the data.