Counting on the most reliable pollworkers

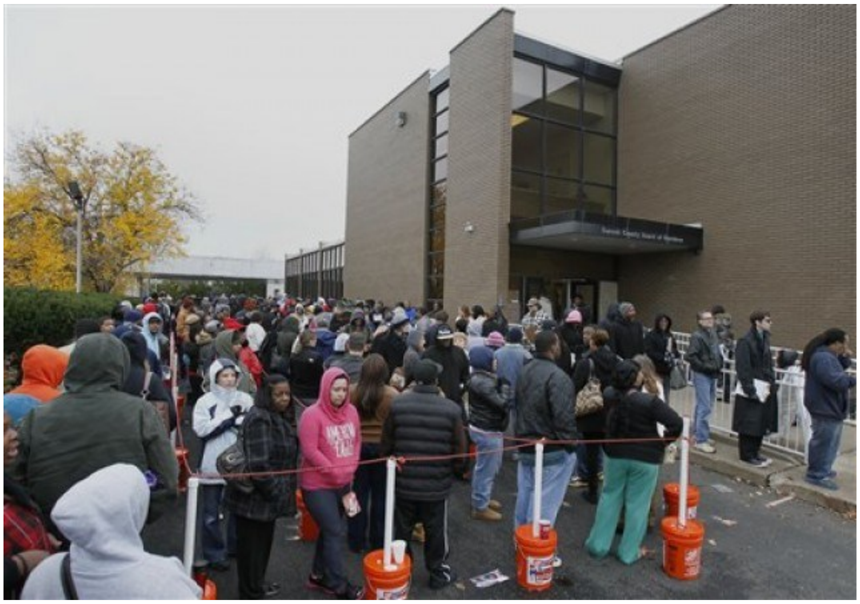
Benjamin Uminsky & Judy Ly

July 12th, 2018

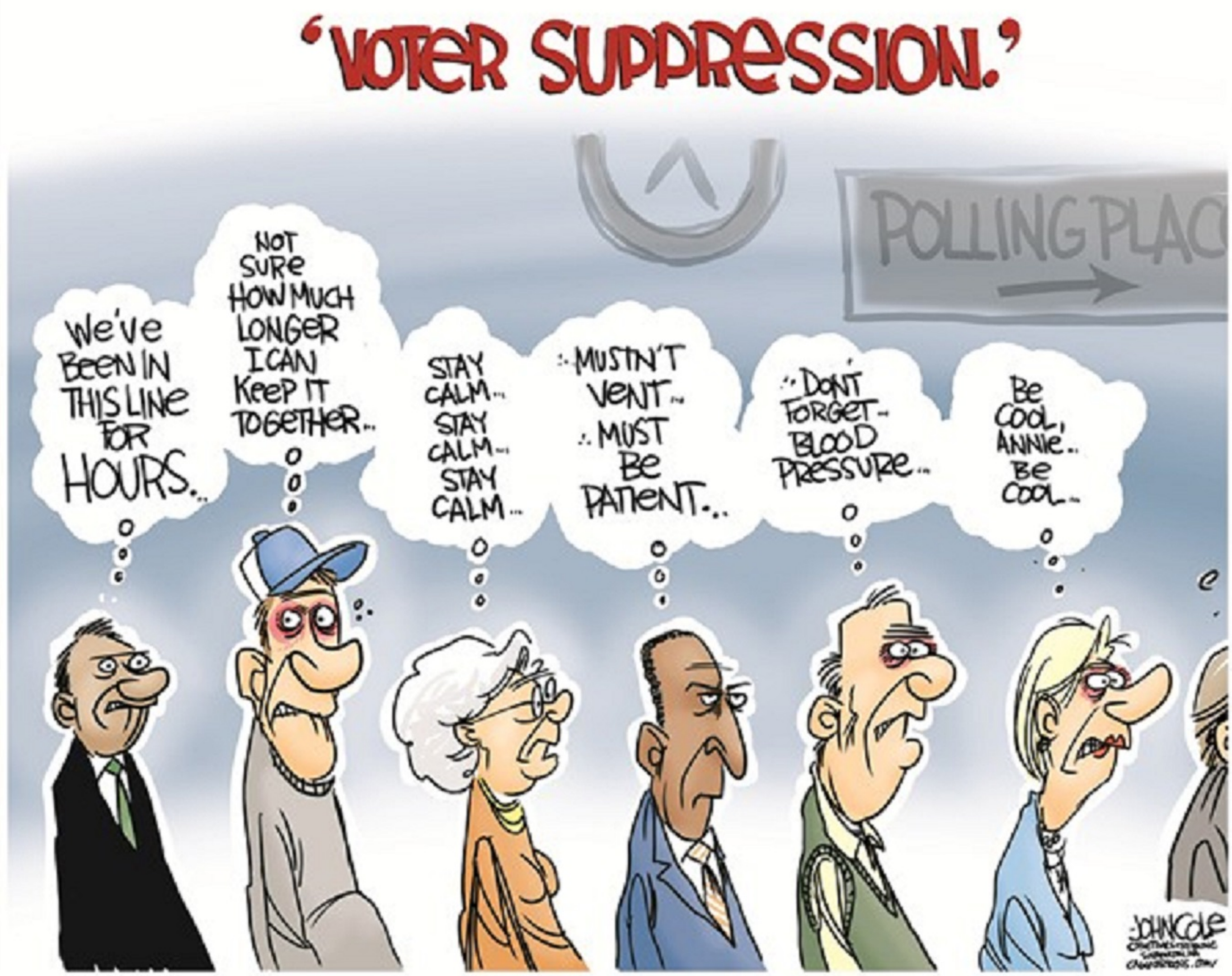
Poll Worker (PW) Recruitment

- What happens when we can’t recruit enough PW’s
- Or cancel last minute
- Or even worse… no show on election day

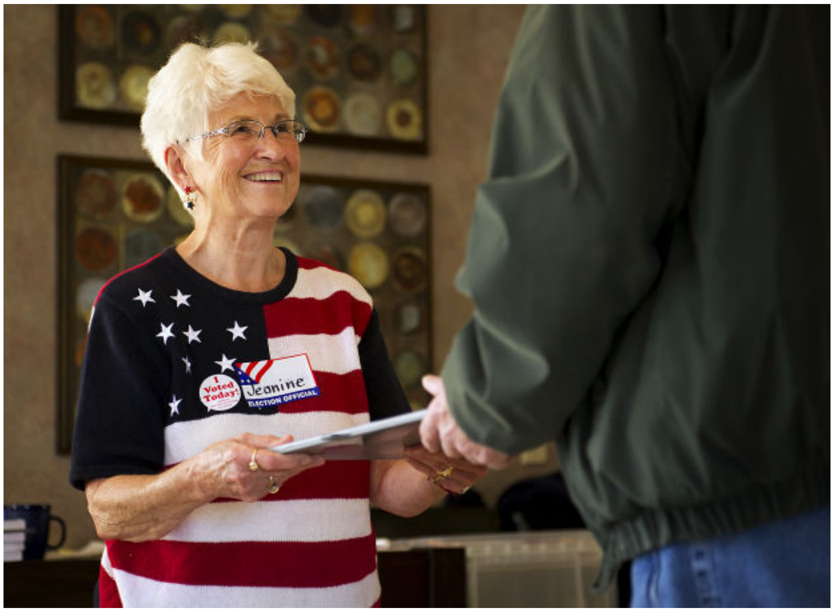
We may end up with…



Which may lead to…



Optimal Selection



So how do we find the reliable pollworker candidates?

ROOM FOR IMPROVEMENT

- Given success rates of 67% for our November 2014 election and 78% for our June 2014 election, clearly room for improvement (less so for our primary elections)
- Can begin better differentiating between commited versus uncommitted PW's
- How do we model the very intuitions that our recruiters are attempting to access
- And find additional patterns that are derived mathematically

GOAL OF THIS PROJECT

- Minimize negative outcomes (PW said no thanks, cancelled last minute, no showed on election day) while still identifying sufficient numbers of PW candidates to actually work
- Costs of incorrectly identifying a negative outcome as a positive outcome: potential late openings of certain polls, last minute shuffling of resources, many more phone calls required

What do we know about our pollworkers

- Some demographic data (gender, age)
- Voting history= civic participation?
- Previous pollworker history predicts future pollworker participation?
- Distance travelled from home to poll site matters?
- Specific outcomes (response)
- PW’s can be split into 6 distinct clusters using unsupervised K means clustering

What is the prediction algorithm doing?

- For each record being predicted on, assigning a class probability (between 0 and 1) to the outcome variable
- Under normal circumstances, anything with a probability >.5 will be predicted as a successful outcome ("A"), anything less will be considered an unsuccessful outcome ("U")
- Setting the threshold higher makes the algorithm more discriminating in its predictions, allowing us to shed our unsuccessful outcomes
- Ended up using a mix of GBM and ADA Boost algorithms

June 2016 Primary

```
##
##     A     C     I     L     O     U
## 22503  3404    75   117  3925   368
```

- We saw a recruitment success rate of 73.6%
- No show rate of nearly 13%
- Cancellation rate of 11.1%

Prediction Algorithm Performance for June 2016 Primary

```
##
##    A    C    I    L    O    U
## 6009  628    9   42  281  110
```

- Algorithm generated a list 25,060 high probability candidates
- Only 7,145 received assignments
- Excellent recruitment success rate of 85%
- Super low no show rate of 3.9%
- Low cancellation rate of 8.8%

November 2016 General

```
##
##     A     B     C     D     E     I     L     N     O     P     R     T
## 27260    28  2371    14     1  4382    68    11  2450   217     4    77
##     U     X
##  1085    43
```

- We saw a recruitment success rate of 71.7%
- No show rate of nearly 6.4%
- Cancellation rate of 6.2%

Prediction Algorithm Performance for November 2016 General

```
##
##    A    C    I    L    O    U
## 9440  561  700   24  215  356
```

- Algorithm generated a list 33,193 high probability candidates
- 11,296 received assignments
- Excellent recruitment success rate of 83.6%
- Low no show rate of 1.9%
- Low cancellation rate of 4.9%

Contact Info

Benjamin Uminsky
Project Manager, Business Intelligence Competency Center (BICC)
Los Angeles County Dept of Health Services
Office of Planning and Analytics
313 N. Figueroa St.
Los Angeles, CA 90012
(**E**) buminsky@dhs.lacounty.gov

Judy Ly
Data Scientist
Los Angeles County Dept of Registrar-Recorder/County Clerk
Executive Office: Data Analytics
12400 Imperial Highway
Norwalk, CA 90650
(**E**) jly@rrcc.lacounty.gov