

COM307 Project 2 Write-up: Wrongful Conviction Data

Khe (Kate) Le

Due date: Oct 12

1. My data

- Exonerated convictions: publicspreadsheet.xlsx
- Non-Exonerated convictions: 2018CrimeStats.csv
- I have decades of exoneration data and only 1 year of conviction data, so I decided to assume that all convictions in the 2018CrimeStats are not exonerated. This helped simplify my calculation and make an insignificant impact to the resulting model. Under my assumption, there are now 11465 non-exonerated convictions and 2515 exonerated convictions thus the ratio of non-exonerated convictions to exonerated convictions is approximately 5:1.
- Hence, my total number of convictions is $11465 + 2515 = 13980$

1. The variables/fields I chose:

- I chose *offense*, *state*, and *race*, because these 3 fields are in both datasets and I'm also interested in them
- In exoneration dataset, these variables are labeled as:

Worst Crime Display - State - Race

- In non-exoneration dataset, these variables are labeled as:

OFFGUIDE - DISTRICT - MONRACE

2. My computation

- In main.py, I came up with two classes Exoneration and NonExoneration.

- Each class has a method for the user to plug in the values of offense, location (state), and race of a conviction to determine the probability that this conviction is exonerated or not.
- My formula for calculating the probability of a conviction being exonerated:

$$p(\text{exonerated}|\text{all that stuff}) = p(\text{exonerated}) * p(\text{thing 1} | \text{exonerated}) * p(\text{thing 2} | \text{exonerated}) * p(\text{thing 3} | \text{exonerated})$$

- My formula for calculating the probability of a conviction being NOT exonerated:

$$p(\text{not exonerated}|\text{all that stuff}) = p(\text{not exonerated}) * p(\text{thing 1} | \text{not exonerated}) * p(\text{thing 2} | \text{not exonerated}) * p(\text{thing 3} | \text{not exonerated})$$

- To compute $p(\text{exonerated})$, I divide the number of exonerations (= 2515) by the number of total convictions (= 13980), thus $p(\text{exonerated}) = 2515/13980 \approx 0.18$
- To compute $p(\text{not exonerated})$, I subtract $p(\text{exonerated})$ from 1, thus $p(\text{not exonerated}) = 1 - 0.18 = 0.82$
- To compute $p(\text{thing 1} | \text{exonerated})$, I compute the probability of each value of thing 1. For example, let thing 1 be race, then the possible values of thing 1 are *white*, *black*, and *asian*. Then I may come up with a result that $p(\text{white}) = 0.5$, $p(\text{black}) = 0.3$, $p(\text{asian}) = 0.2$. I used the same calculation for other fields (state, offense) and convictions that are not exonerated.
- To view the computed probability of each field and their values in both dataset, see my two log files ***exnVarsProb.log*** and ***nonexnVarsProb.log***