Predicting Arrest in Chicago Sexual Assault Cases

• • •

By Heather Robbins

Sexual Assault In A Big City: Is There A Lack Of Priority?

March 26, 2015 1:45 PM By Mason Johnson

170 People Stand In Silence Against Sexual Assault

April 28, 2017 4:29 PM

Filed Under: Daley Plaza, Lisa Fielding, rape, Rape Victim Advocates, sexual assault, Silence

CHICAGO (CBS) — An unusual sight at Daley Plaza in Chicago during the lunch hour Friday as 170 men and women stood in silence in front of the 50-foot Picasso. WBBM's Lisa Fielding reports.

SLUTWALK CHICAGO AUGUST 20TH 2016

Why aren't we taking rape seriously?

by Ashe Schow | Apr 2, 2015, 5:00 AM

END RAPE CULTURE
SUPPORT THE VICTIMS
END RAPE

UNIVERSITY OF CHICAGO AMONG COLLEGES INVESTIGATED FOR MISHANDLING CAMPUS SEXUAL ASSAULT CASES
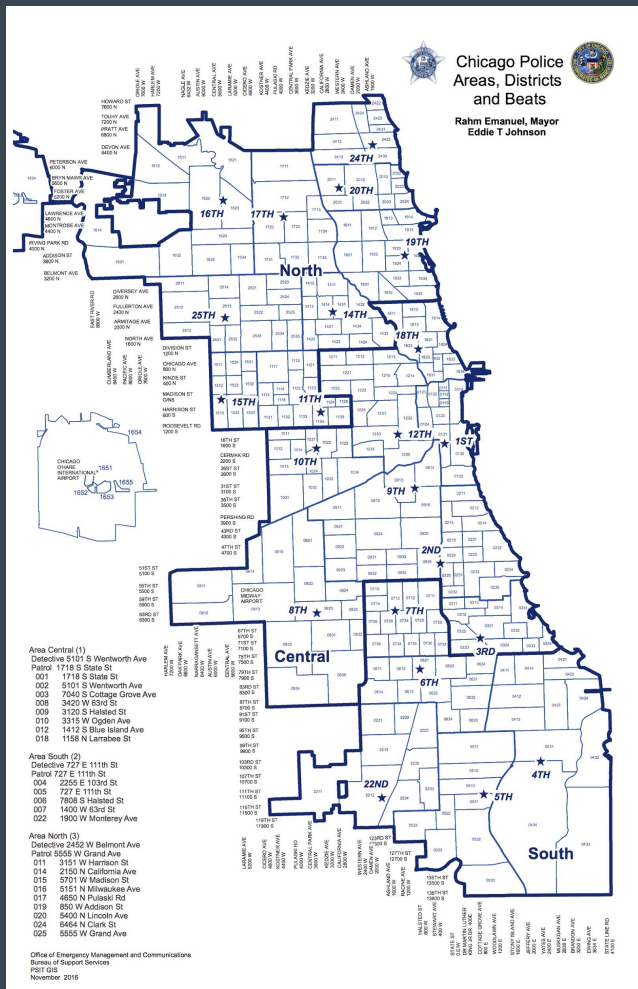
Federal authorities investigating sexual violence complaints at several area colleges

Can a model predict if an arrest will be made when sexual assault is reported in Chicago?

Data Acquisition

City of Chicago Data Portal:
- Crimes 2001-Present
- Police Station Locations

College/University Data:

- Criteria:
    - Must have over 3500 students
    - Must have residence halls

- Colleges/Universities Included:
    - Loyola University
    - Northeastern Illinois University
    - DePaul University
    - University of Illinois at Chicago (East & West)
    - Illinois Institute of Technology
    - Columbia College
    - School of The Art Institute Chicago
    - University of Chicago
    - Chicago State University

- Locations:
   - Addresses chosen are central to the residence halls

Data:

- Location:
  - Latitude & Longitude of crime scene
  - Beat, Community Area, District, Ward
  - General Description of Location (Residence, Public Property, Vehicle Place of Business, etc)
  - Distance from Nearest Campus and Distance from Nearest Police Station
- Date/Time:
  - Year, Month, Day of the Month, Day of the Week, Hour of the Day, Time of Day (Early Morning, Morning, Afternoon, Evening)
- Domestic
  - If it was related to a domestic dispute/domestic violence
- Type of Assault:
  - Aggravated, non-aggravated, predatory, attempted-aggravated and attempted-nonaggravated

Modeling: Random Forest Classifier

Round 1: 2001-2016

Rows/Reports: 22, 450

	Importance
Nearest Campus	0.237101
Nearest Police Station	0.223773
NON-AGGRAVATED	0.084154
Residential Property	0.036227
Domestic	0.027212
PREDATORY	0.026917
Early Morning	0.023382
Evening	0.023236
2016	0.022270
AGGRAVATED	0.021523
Public	0.020481

Feature importances
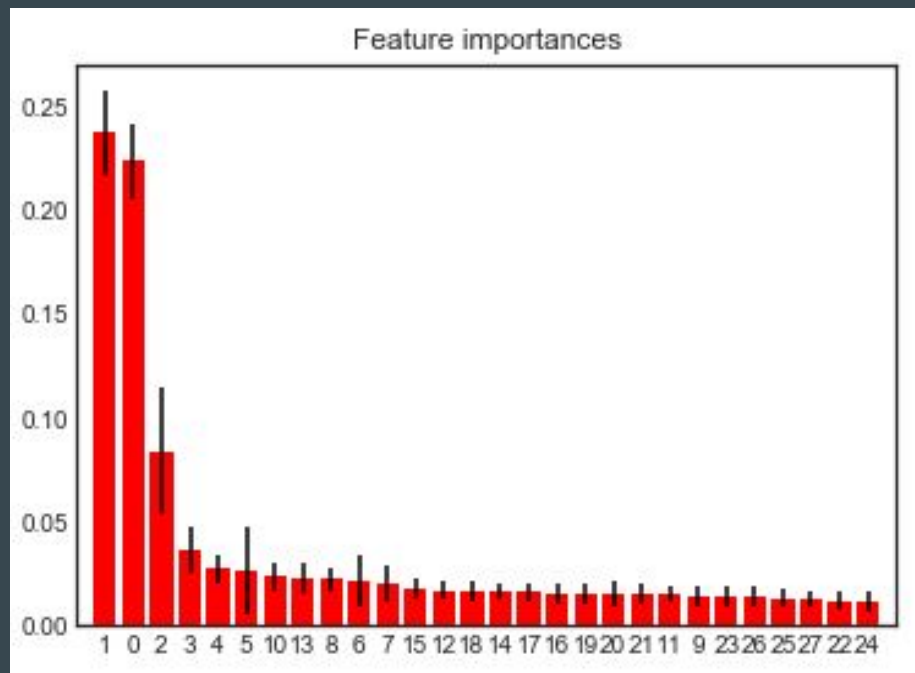
Tuning:

- Before tuning:
  - Mean Accuracy: 0.832, but was classifying everything as a non-arrest
- Used Grid Search and set to optimize Recall scores
- After Tuning:
  - Mean Accuracy: 0.670, but a much better Recall score

	predicted_no_arrest	predicted_arrest
no_arrest_actual	3145	1596
arrest_actual	437	510

	precision	recall	f1-score	support
0	0.89	0.69	0.78	4769
1	0.26	0.55	0.35	919
avg / total	0.79	0.67	0.71	5688

Round 2: 2017 Data

Rows/Reports: 403

Mean Accuracy: 0.720

	predicted_no_arrest	predicted_arrest
no_arrest_actual	285	109
arrest_actual	4	5

	precision	recall	f1-score	support
0	0.99	0.72	0.83	394
1	0.04	0.56	0.08	9
avg / total	0.97	0.72	0.82	403

Random Forest Classifier ROC Curve

ROC curve (area = 0.696)
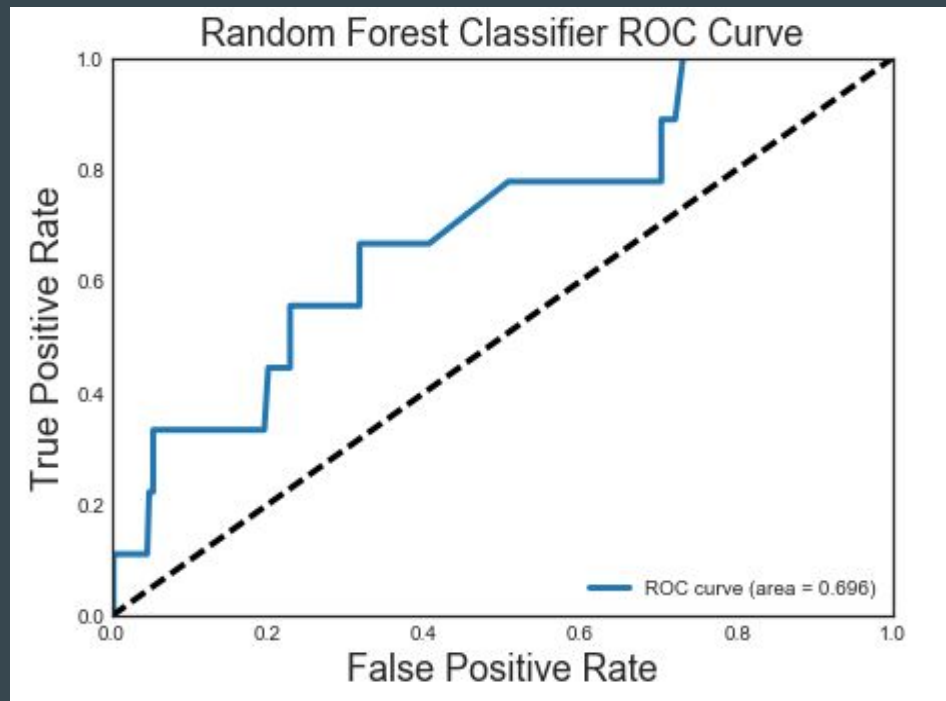
How Can This Data Be Used?:

- This model/information is just one small piece of a much bigger puzzle, so this isn't intended to be immediately actionable.
    - EXCEPT:
        - Keep the conversation going
        - Chi Hack Night

Future Steps/Improvements:

- Try to find more data:
    - Demographics
    - How each geographical area is staffed
    - Police training & salaries
    - Amount of time between incident and filing a report
    - Overall crime
    - Population density
- Look more closely at the relationship between distance from campuses and distance from police stations

THANK YOU!

Richard, Linda & Chris

Fellow classmates

Chicago Data Portal

The Data Gods

Frontlines