IBM Qiskit Hackathon

MSiA Group 13

Isaac Choi Sara Ho Tiffany Ke Matt Ko Louis-Charles Genereux

The team chose to focus on identifying factors that make cities more susceptible of seeing police shootings

Key question

Can we identify factors that make certain cities at a higher risk of suffering police shooting catastrophes?

Building confidence around the answer to this question could have an impact on public policy, police training, education and beyond.

Approach

The team decided to study police shootings on a per capita basis across different US cities or regions.

After summarising shootings at the city level (and then county), we gathered socio-economic (education, poverty, unemployment, GDP per capita) indicators and police statistics (funding, number of officers per capita, number of complaints).

We first explored the strength of correlation of these variables to shootings through tree methods then turned to Qiskit.

Data Processing

Dataset

- Final dataset: combination of 4 datasets (Kaggle + external sources*)
- City and County data bridged using Google Maps API
- Kaggle dataset was aggregated at city level, then calculated into percentages for each feature

Variables

- Feature Scaling: MinMaxScaler scales features to [0, 1] (good for SVM)
- Target variable: **level of police killings per 10k arrests**
- Predictors

* Socio-economic (GDP, urbanization, employment, education)
  Police-department stats (officers per capita, complaints per capita, budget per capita)

Feature Selection - Shootings data

% of killed who carried gun
% of killed who were unarmed
% of killed who were minority
% of killed who had sign of mental illness
% of killed who attacked
% of killed who were fleeing
% of killed age 10 and below
% of killed 30 and below
% of killed 50 and below
% of killed who were male
% of killed whose threat undertermined
% of killed whose killer had body camera on

- Aggregated Shootings data
- Every column is a percentage
- Some categorical variables are left out on purpose to avoid multicollinearity

Feature Selection - Socioeconomic data

Estimated percent of people of all ages in poverty 2019
Estimated percent of people age 0-17 in poverty 2019
Estimated percent of related children age 5-17 in families in poverty 2019
Estimate of median household income 2019
Unemployment_rate_2019

- Poverty / Income variables
- These variables may indicate economic status of the city

POP_ESTIMATE_2019
white_population

- Overall Demographic variables

Feature Selection - Education levels

Percent of adults with less than a high school diploma 2015-19
Percent of adults with a high school diploma only 2015-19
Percent of adults completing some college or associate's degree 2015-19
Percent of adults with a bachelor's degree or higher 2015-19

- Education variables
- May indicate education level of adults in each city
- These variables are averages of percentages from 2015 to 2019

Feature Selection - Police Agency data

calc_officers_per_10k_population
calc_fines_forfeitures_per_resident
calc_police_spending_per_resident

- Police presence variables
- These variables may indicate how active and involved the police is in each city

Final Features for Qiskit

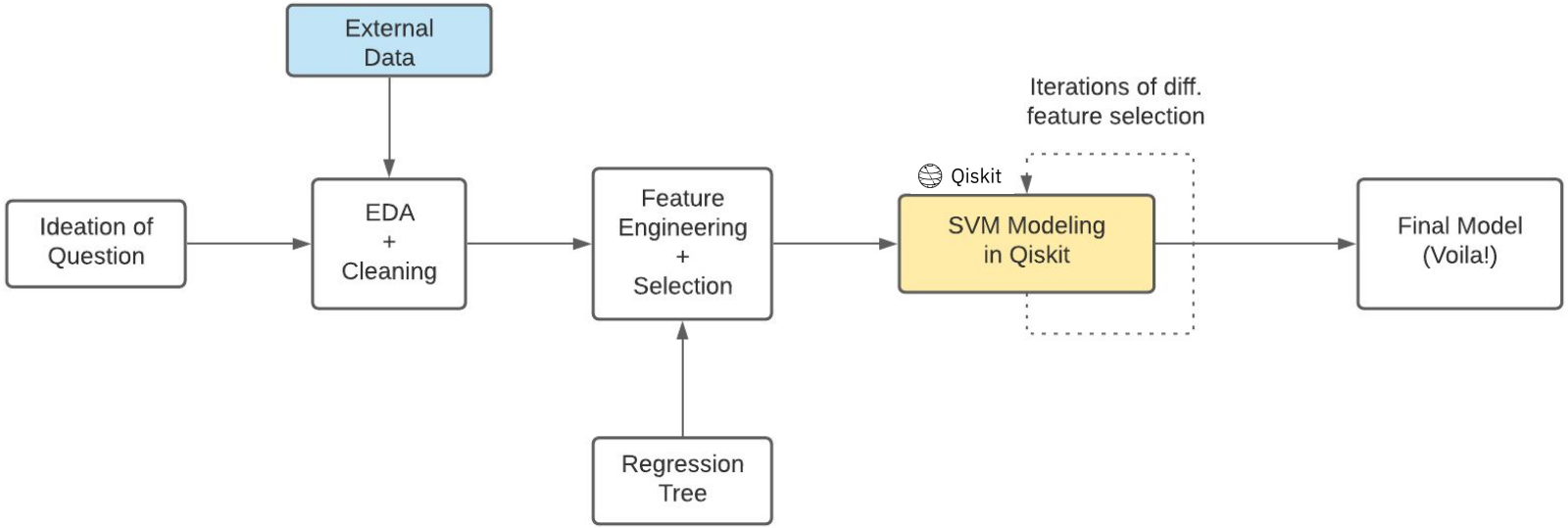
- We had to limit number of features for effective use of Qiskit and the quantum computer
- Only Subsets from shootings were selected
- All Education variables were included. While there are many hypothesis involving economics and race with police killings, we wanted to see if education level in a city plays any role in police killings
- All police presence variables included
- Percentage of white population in city included

% of killed who carried gun
% of killed who were unarmed

Percent of adults with less than a high school diploma 2015-19
Percent of adults with a high school diploma only 2015-19
Percent of adults completing some college or associate's degree 2015-19
Percent of adults with a bachelor's degree or higher 2015-19

calc_officers_per_10k_population
calc_fines_forfeitures_per_resident
calc_police_spending_per_resident
white_population

Method and Process



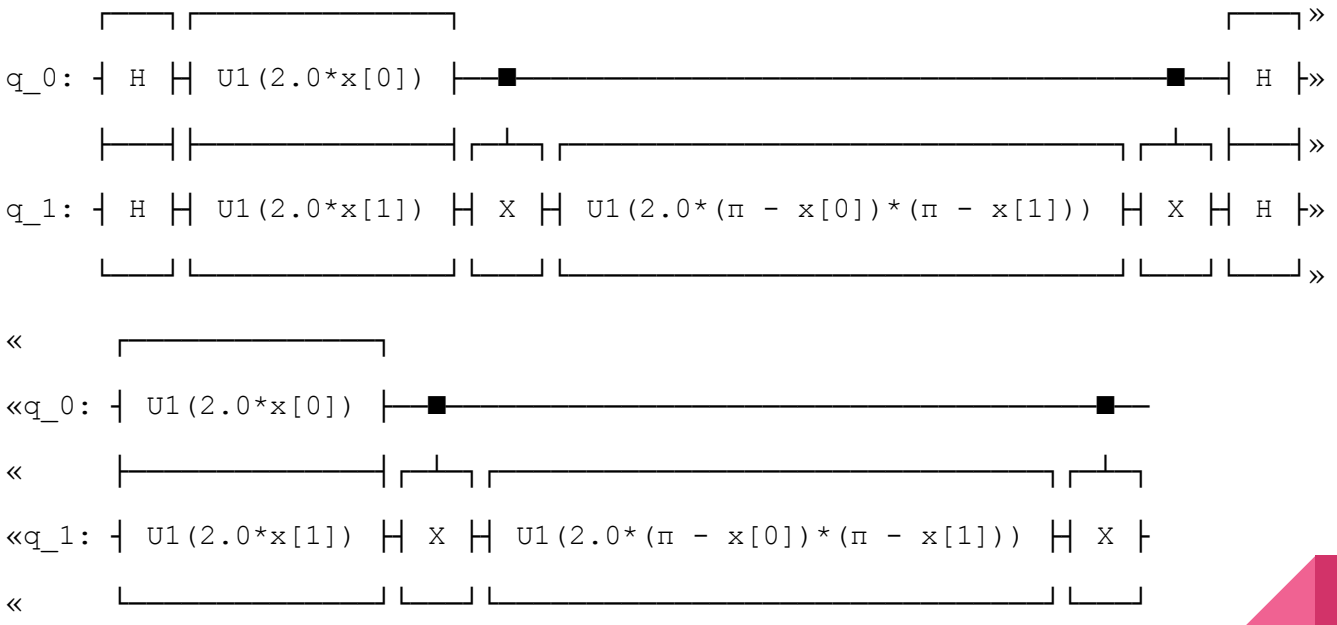
Feature Map

$$V(\Phi(\vec{x})) = U(\Phi(\vec{x})) \otimes H^{n} \otimes U(\Phi(\vec{x})) \otimes H^{n}$$

- According to medium article, this is the final feature for quantum SVM
- Fortunately for us Qiskit included this feature map in its aqua.algorithm package called QSVM

*https://medium.com/@patrick.huembeli/introduction-into-quantum-support-vector-machines-727f3ccfa2b4

Qiskit


q_0: H U1(2.0*x[0]) H
q_1: H U1(2.0*x[1]) X U1(2.0*(π - x[0])*(π - x[1])) X H
«q_0: U1(2.0*x[0])
«q_1: U1(2.0*x[1]) X U1(2.0*(π - x[0])*(π - x[1])) X

Results from Qiskit

- We first had to limit number of columns and number of rows
- We chose backend to be ibmq_qasm_simulator due to long queues of other systems.
- The accuracy of the model was at 86%

```
print(result['testing_accuracy'])
print(result['test_success_ratio'])
```

```
0.8571428571428571
0.8571428571428571
```

In closing, our data exploration process revealed just how difficult the problem of police shootings is to crack for public leaders

Studying events of police shootings ex-post creates the risk of making claims based on spurious relationships

When running regression trees on police deaths per capita, we identified the following as highly significant:

- County education levels
- County GDP per capita
- Number of historical PD complaints per capita
- Officers per capita

While these variables **correlate** highly with shootings, one cannot claim that they **cause** them**.** These signals might even be the result of violence

Low data availability and questionable quality might limit policy makers' confidence in model findings

We found that data on police departments are often crowd-sourced and of varying levels of exhaustiveness / quality based on the geography

Thankfully, police shootings do not occur at the same pace as other events typically modelled. This however, affects the availability of "training data", limiting prediction accuracy

Finally, and most importantly, this issue involves human behavior and will not be solved through quantitative methods alone

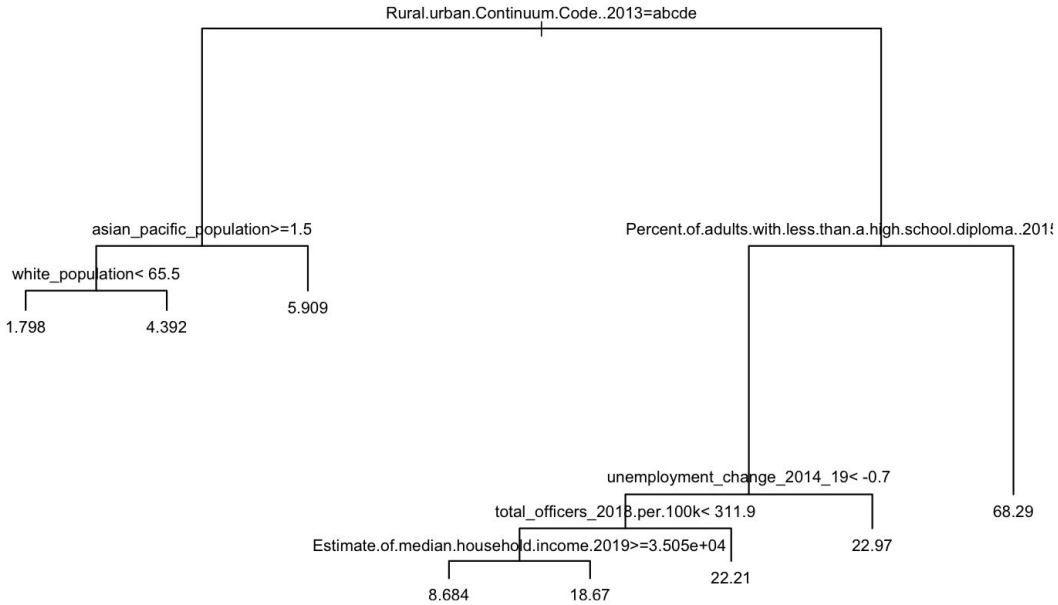
Human behavior can be illogical, so even in a world of perfect data availability, unavoidable catastrophes will occur

Our initial data exploration and analysis uncovered mental-health, socio-economic and racial phenomena that have showed correlations to police shootings.

These are important social issues that should be addressed regardless of model findings.

Initial data exploration

The impact of demographic factors on police shootings per capita



Initial findings:

- Higher urbanization correlates with lower occurrences of shootings
- Higher educational attainment correlates with lower occurrences of shootings
- Some demographic differences exist between counties experiencing varying levels of occurrences

These relationships not equate to causation

More analysis required to determine correlation across these predictor variables

Initial data exploration

Adding police statistics to demographic data as predictors of shootings

Initial findings:

- Higher number of complaints correlates with police violence

Complaints may be leading indicator of eventual future shootings or may also be in response to violence; making it difficult to establish causation