

# Predicting Crime

## Midterm Report

Clare Xu (V00812505)  
Richard Lui (V00221256)

Nan Yang (V00811300)  
Meredith Olsen-Maier (V00181696)

### Problem Description & Inspirations

For this project, we are looking to predict what category of crime is likely to happen in Victoria and Saanich during certain time.

With many crimes happening in Greater Victoria each month, it can be hard to identify where in Greater Victoria it happens in and what category of crime the crime falls into. With this project, we hope to give policy makers and police additional research when assisting the public. We also hope to inform the public about particular crimes that happen in certain times of the day, week, month, and year, so the public can better prepare themselves against crimes.

### Data & Our Prediction

#### Input:

At first we were going to use both the data from <https://moto.data.socrata.com/> [1] and the data which we obtain in person from each municipality, since the data from the website only cover crimes in Victoria and Central Saanich spanning from 2006 to 2016. However, due to the Freedom of Information Act, it might take 4 to 6 weeks to gather the data in person, so we decided to only use the data from the website.

The original dataset from the website has data on the category of a crime, the time of the day, the day of the week, the month, and the year spanning

back to 2006. We limited our timeline from 2011 to now. Additionally, in order to clean up the incidents that we do not consider crimes such as traffic tickets, car accidents and noise violations, we grouped the data based on their type and deleted parts we tend to ignore as crimes in Excel before actually running program on it.

Our data is stored in a .csv that is loaded into our program, and then converted into a  $9 \times n$  matrix, where  $n$  is the number of rows and 9 is the number of columns. The columns contain data such as: incident type (primary and parent), address, city, state, location (latitude and longitude), hour of the day, day of the week, month of the year, day of the month, and year.

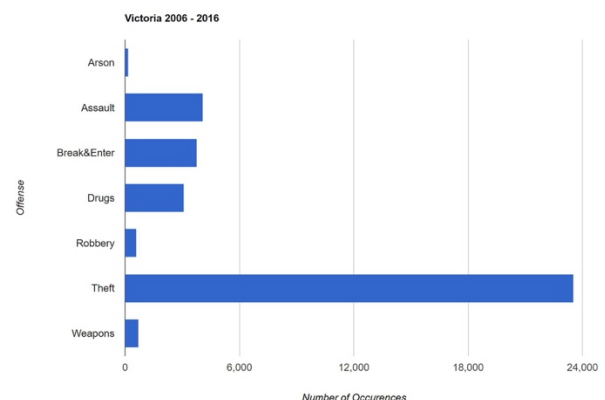


Figure 1. Graph of offense type and number of occurrence in Victoria from 2006 to 2016.

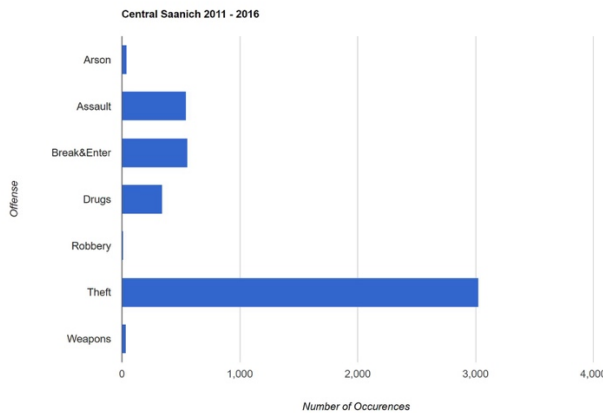


Figure 2. Graph of offense type and number of occurrence in Central Saanich from 2011 to 2016.

### Output:

Our output will be the predicted category of crime, given the time of day, and day of the week.

Running the cv.py file [2] from our lab on a very small randomized segment of our data we found the best results were given using sklearn's Perceptron classifier:

GaussianNB:	0.320000
DecisionTreeClassifier:	0.320000
MultinomialNB:	0.320000
BernoulliNB:	0.480000
Perceptron:	0.520000
LogisticRegression:	0.400000

### Our Plan for the Next Stage

Moving forward with this project, we plan to finish training our data and quickly move to implementing our classifying algorithm. AS previously stated the Perceptron classifier performed the best on a small subset, however due to the text format of our dataset we will be sticking to our original plan and implementing a Naïve Bayes classifier.

### Updated Timeline

OCT. 30 - NOV. 5

Training data.

NOV. 6 - 12

Algorithm writing.

NOV.13 - 19

Test the algorithm with dataset, and analysis solution.

NOV. 20 - 26

Evaluation and report writing.

NOV. 27 - DEC 3

Final Presentation.

### Workload Distribution

Member	Distribution
Clare	Training data, analysis solution, evaluation and report writing, final presentation
Meredith	Training data, analysis solution, evaluation and report writing, final presentation
Richard	Algorithm writing, test algorithm with dataset, analysis solution, final presentation.
Nan	Algorithm writing, test algorithm with dataset, analysis solution, final presentation.

### References

- [1] Socrata, "Victoria(BC) Police Department", and "Central Saanich Police Service", 2016, <https://moto.data.socrata.com/>.
- [2] C. Shortt, "cv.py", SENG 474 Lab 04, University of Victoria, 2016.