

- **Progress so far**

As a recap, my final project will serve as a prototype for my master project. It focuses on improving the accuracy of crime prediction based on current models. I have already completed the baseline model, which does not use any specific model. Instead, it relies on historical data to predict the next possible hotspots.

I have compared different time frames to determine the date range that provides the best efficiency. The results indicate that the larger the time frame, the more efficient the Predictive Efficiency Index (PEI) becomes. For example, initially using a 14-day time frame yielded a PEI of only 0.5. However, as I increased the time frame, the PEI also increased. Analyzing data from the calendar year 2012, specifically between March and December, I found that a 150-day time frame produced the best result. The PEI was close to 0.967, indicating highly accurate hot spot predictions. Almost all hot spots were included in this prediction. Table 1 as shown in appendix shows a increase of PEI as time frame increased.

Data: [Dataset](#)

- **Challenges**

At present, I've initiated the training dataset with only street crime, comprising a total of 20,000 cases. If I were to add more yearly data, the program runtime would increase significantly. Furthermore, upon visualizing the data, I observed a considerable number of empty cells, indicating a significant imbalance in the dataset. Unfortunately, due to the correlation between coordinates and cells, eliminating all empty cells is not feasible.

For a large dataset, based on my research, I have a few options. I can either implement data sampling, which may lower accuracy (similar to what I am currently doing), explore various methods, or leverage cloud computing for parallel processing.

Regarding class imbalance, I acknowledge the need for further research to gain a better understanding. Currently, I believe data transformation could prove beneficial in reducing the number of empty values.

- **Collaboration** (if applicable)
 - This is not applicable as I am working on this project alone
- **Next steps**

Moving forward, my plan is to continue working on the program. Using the baseline model as a reference, I intend to implement a Convolutional Neural Network (CNN) model to analyze the data. The goal is to compare the effectiveness of the CNN model against the baseline.

Since I haven't worked with any models before, the implementation will pose a challenge, and there will be a learning curve involved. Nevertheless, I am prepared to navigate through this learning process as I strive to enhance the capabilities of the model.

Table 1. 2012 Street Crime Prediction with different time frame

Time Frame = 14	Time Frame = 30	Time Frame = 90	Time Frame = 150
<p>Layer 1 - Time Interval: 2012-03-01 to 2012-03-15 Total count for the actual top 100 cells: 299.0 Sum of Points: 910 points</p>	<p>Layer 1 - Time Interval: 2012-03-01 to 2012-03-31 Total count for the actual top 100 cells: 524.0 Sum of Points: 1941 points</p>	<p>Layer 1 - Time Interval: 2012-03-01 to 2012-05-30 Total count for the actual top 100 cells: 1498.0 Sum of Points: 6429 points</p>	<p>Layer 1 - Time Interval: 2012-03-01 to 2012-07-29 Total count for the actual top 100 cells: 2537.0 Sum of Points: 11440 points</p>
<p>Layer 2 - Time Interval: 2012-03-15 to 2012-03-29 Total count using indices from the last layer: 144.0 Total count for the actual top 100 cells: 285.0 Sum of Points: 920 points PEI: 0.5052631578947369</p>	<p>Layer 2 - Time Interval: 2012-03-31 to 2012-04-30 Total count using indices from the last layer: 398.0 Total count for the actual top 100 cells: 586.0 Sum of Points: 2215 points PEI: 0.6791808873720137</p>	<p>Layer 2 - Time Interval: 2012-05-30 to 2012-08-28 Total count using indices from the last layer: 1454.0 Total count for the actual top 100 cells: 1673.0 Sum of Points: 7581 points PEI: 0.8690974297668859</p>	<p>Layer 2 - Time Interval: 2012-07-29 to 2012-12-26 Total count using indices from the last layer: 2390.0 Total count for the actual top 100 cells: 2633.0 Sum of Points: 11621 points PEI: 0.9077098366881884</p>
<p>Layer 3 - Time Interval: 2012-03-29 to 2012-04-12 Total count using indices from the last layer: 153.0 Total count for the actual top 100 cells: 298.0 Sum of Points: 904 points PEI: 0.5134228187919463</p>	<p>Layer 3 - Time Interval: 2012-04-30 to 2012-05-30 Total count using indices from the last layer: 411.0 Total count for the actual top 100 cells: 583.0 Sum of Points: 2273 points PEI: 0.7049742710120069</p>	<p>Layer 3 - Time Interval: 2012-08-28 to 2012-11-26 Total count using indices from the last layer: 1821.0 Total count for the actual top 100 cells: 2102.0 Sum of Points: 9051 points PEI: 0.8663177925784968</p>	
...	...		
...	...		

<p>Layer 21 - Time Interval: 2012-12-06 to 2012-12-20</p> <p>Total count using indices from the last layer: 250.0</p> <p>Total count for the actual top 100 cells: 471.0</p> <p>Sum of Points: 1693 points</p> <p>PEI: 0.5307855626326965</p>	<p>Layer 10 - Time Interval: 2012-11-26 to 2012-12-26</p> <p>Total count using indices from the last layer: 451.0</p> <p>Total count for the actual top 100 cells: 617.0</p> <p>Sum of Points: 2332 points</p> <p>PEI: 0.7309562398703405</p>		
---	---	--	--