Literature Review: Ethical and Methodological Challenges in Predictive Policing

Problem Definition

Predictive policing is a data-led proactive approach to policing, where analysis and data models are used in order to predict people and areas sensitive to crime so more resources can be allocated to them (Gstrein et al., 2019; Hung & Yen, 2020). Despite the incredible potential of predictive policing, it is also home to a plethora of ethical issues, especially involving bias which can affect minorities especially (Almasoud & Idowu, 2024). This is partially because of the nature of past data used in training the models, which carries and reinforces biases, but also because of the way the data is handled and fed to the model (Purves, 2022). This review examines the ethical and methodological dimensions of predictive policing systems, analyzing their potential benefits and drawbacks, and proposing strategies to enhance fairness and accountability.

Project Objectives

In this review, we aim to:

1. Explore some of the many methods employed in predictive policing and see how they work and where they might be room for improvement with regards to upholding ethics and giving people their due rights.
2. Analyze how predictive policing impacts society, in particular the minorities.
3. Evaluate fairness metrics and bias mitigation strategies in existing literature.
4. Make recommendations in order to improve the discussed methodologies from an ethical perspective.

Analysis, Results & Discussion:

Note: Most of the code involved in this case was written with the help of Gemini, since it was done in Google Colab, which comes with Gemini built in.

For the purpose of this analysis, what we care about is that crime occurs, and the location at which it occurs. While the nature of crime can be an interesting variable to use in order to better equip the authorities, it’s not something we take into consideration for this review, since our models are going to be more targeted towards predicting the area in which crime occurs.

We use a more cyclical representation of time using sin & cos, for hours in the day, days in the week, and the month.

In order to make the data more feasible for model training, we had to select variables from given data which were valuable to us, so we dropped 'Case Number', 'IUCR', 'Updated On', 'Year', 'FBI Code', 'Description'.

Then, we started working on managing the qualitative variables better, primarily the “Primary Type” & “Location Details” variables. We started off by reducing the number of unique values by using categorization, which was especially helpful with Location Details, since we managed to bring the number down from hundreds to a mere 9 values.

Once we had this covered, we used hot-encoding to essentially give the two qualitative variables a quantitative representation. For each value in the qualitative variable (which are now much fewer values compared to the start thanks to our categorization work), we create a column and give all the entries a binary value for whether that value exists in an entry or not. Now that we have this, we can use this information to train our models.

We now pick our response variables, which in this case would be 'Longitude', 'Latitude', 'Location', 'X Coordinate', & 'Y Coordinate'.

We also scale the x-variables using a standard scaler.

Now we get to making turning and testing splits. For this case, we’re gonna use a standard 80-20 split.

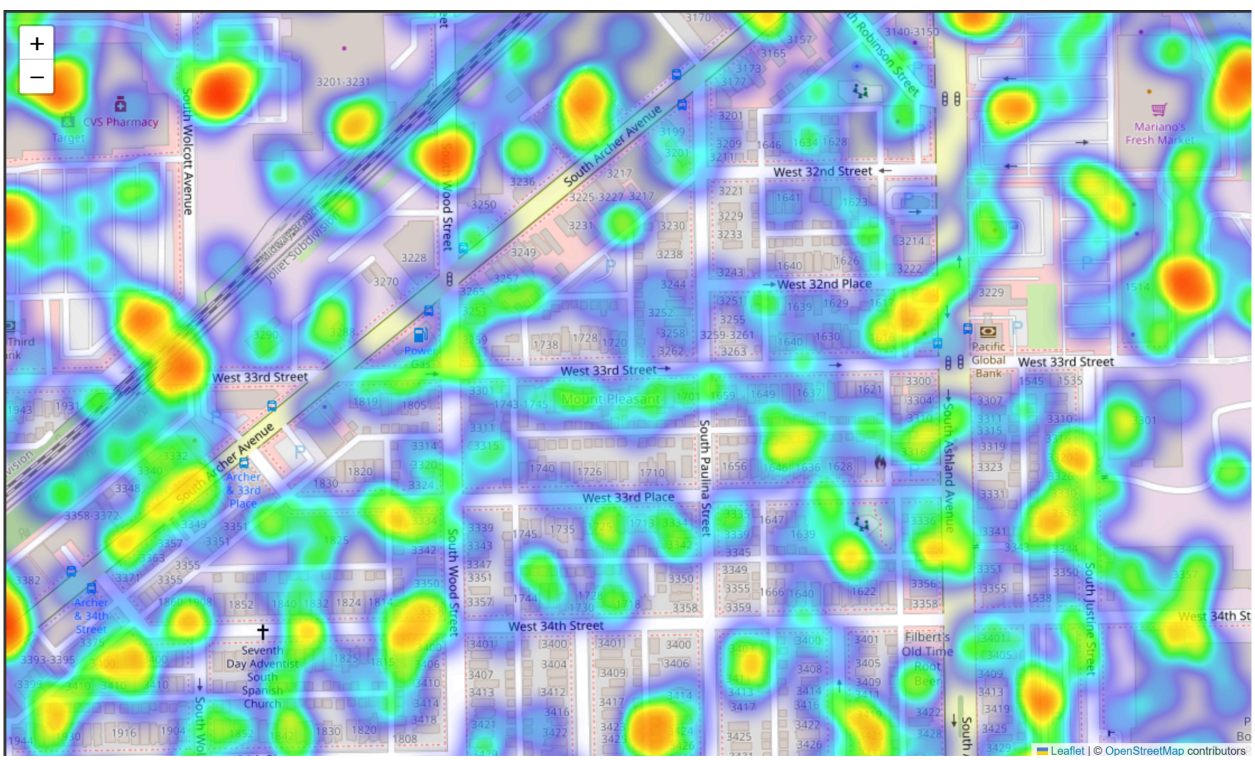
Once our training-testing split is set, we start picking up and training our models. Since we are looking to see zones and points where crimes are likely to occur, we would go for regression or clustering based approaches instead of classification approaches.

The first method we use is the random forest model for regression.

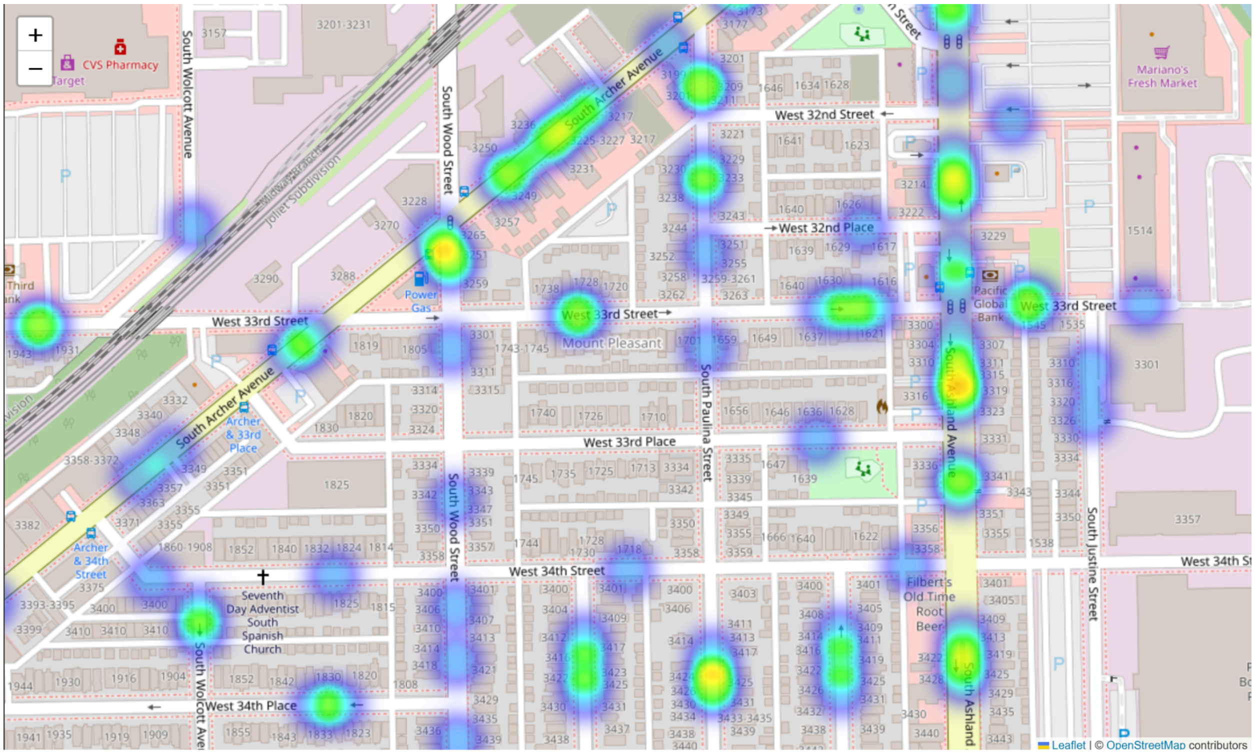
Random Forest Regression:

Random Forest works in a very simple way. First, it creates multiple subsets of the data by sampling with replacement. Then, for each subset, we build a decision tree using a random selection of features from each split. After that, the output is the average of predictions from all trees.

In our case, once the model was run, we used the output to build a heatmap. A screenshot of the heatmap is given below:



Compared to the original data, which gave us this heatmap for the same area:



As we can see, the model has hyperbolized the crime points by a lot. We calculated the mean square error for predictions, and found that for certain blocks, the model gave higher error rates for different neighborhoods:

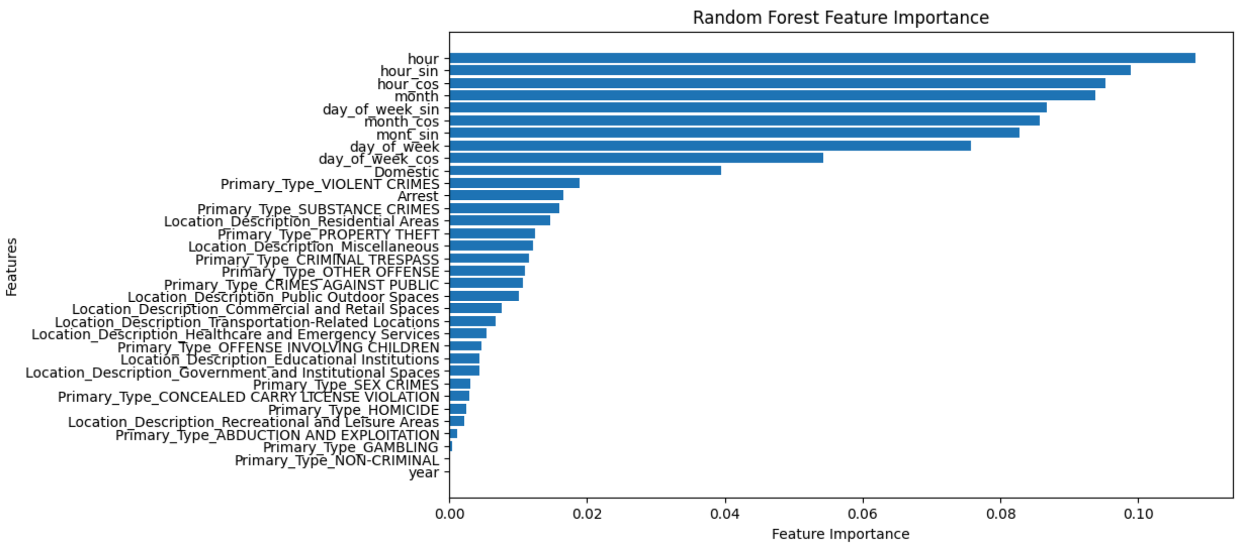
```
Top 10 blocks with highest Longitude MSE:
Block
002XX N LOWER MICHIGAN AVE    0.062500
0000X W B17 ST                0.060211
043XX N MANNHEIM RD           0.040844
0000X W K9 ST                 0.037502
086XX W BERWYN AVE            0.036657
100XX W KENNEDY EXPY IB       0.036561
097XX W FOSTER AVE            0.036068
0000X W B6 ST                 0.034753
001XX E 50TH ST               0.034656
007XX W OHARE ST              0.034098
Name: mean_squared_error, dtype: float64

Top 10 blocks with highest Latitude MSE:
```

Block	mean_squared_error
002XX N LOWER MICHIGAN AVE	0.094368
002XX E 134TH ST	0.052659
129XX S PARNELL AVE	0.051791
120XX S PARNELL AVE	0.049000
064XX N WASHTENAW AVE	0.047009
063XX N LAKEWOOD AVE	0.046923
009XX E 131ST ST	0.046904
132XX S BRAINARD AVE	0.046332
134XX S BRANDON AVE	0.046250
130XX S HOUSTON AVE	0.044878

& weirdly enough, the F1 score tells a different story:

```
F1 Score for Latitude: 0.9788902875956231
F1 Score for Longitude: 0.9999806332913721
```

If we look at the effect each variable has on the output, we see this trend:



The model gives more importance to time of the day than it does to type of location or type of crime. This gives us the idea that the model isn't necessarily predicting where a crime might happen based on where it has happened in the past, but on what time and date it is.
Next, we see how the model & data perform for k-anonymity, l-diversity, t-closeness, reidentification risks and differential privacy stats:

k-Anonymity = 1:

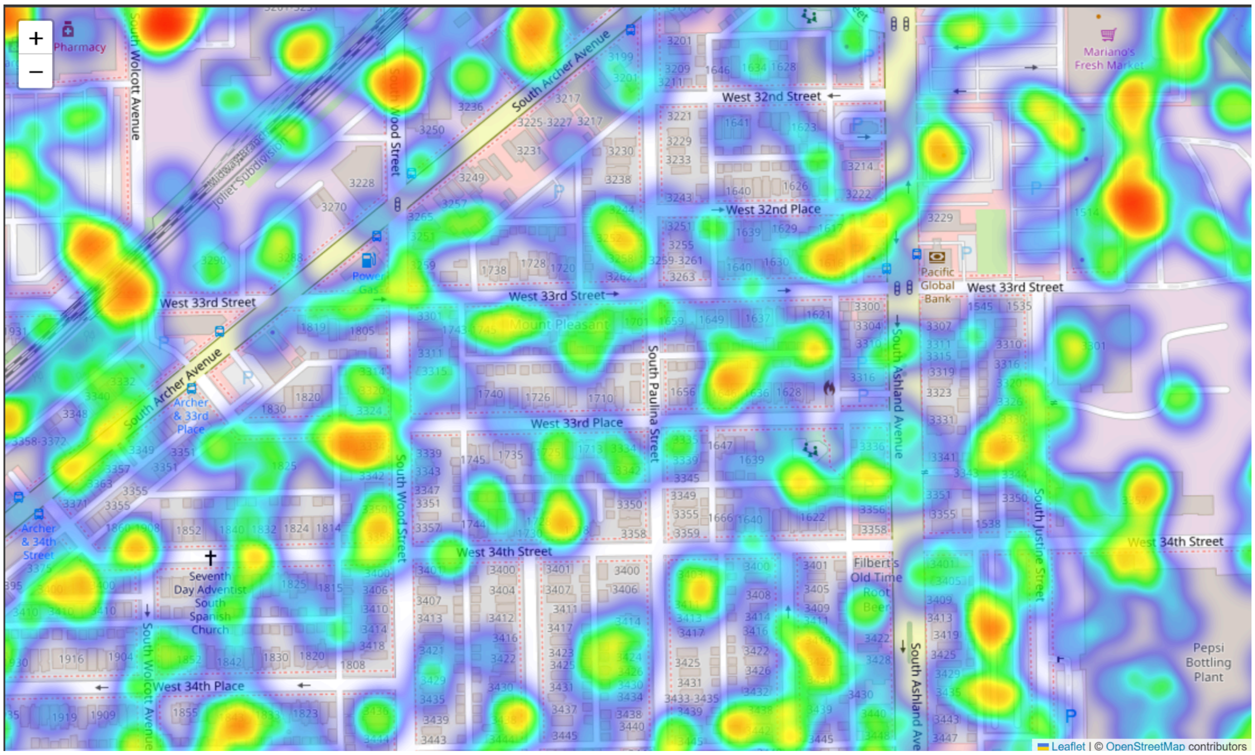
Considering the results above, we can make the conclusion that the model has overfit.The Random Forest model has learned exact patterns from the dataset, making the records easily distinguishable.
This means that the model has little to no generalization and may be exposing specific details of the training data.

l-Diversity = 1:

The sensitive attributes in the dataset (or predictions made by the model) are not well-protected and are highly predictable.
This could mean:
The dataset lacks diversity in sensitive attributes.
The model is overfitting and directly mapping inputs to sensitive outputs.

In order to solve this problem, we re-do the training testing split, this time to a 60-40 split, and see how the model responds.

With a 60-40 split, we got the heatmap as:



The F1 score and other metrics also haven't changed to a satisfactory degree, showing that the model is still being overtrained. The issue with considering time as the major contributor still persists.

Results

The literature reveals mixed outcomes for predictive policing:

1. **Effectiveness**: While systems like PredPol have shown that they can be resourceful when solving allocation problems, their impact on overall crime reduction is contested by many (Hung & Yen, 2020).
2. **Ethical Failures**: Predictive policing allows vulnerable communities to be overpoliced because of the nature of data collection for such efforts (Purves, 2022; Gstrein et al., 2019).

3. **Bias Mitigation Success**: Emerging techniques like Conditional Score Recalibration, show promise in reducing specific biases without sacrificing accuracy, but a lot of effort is required before we can see such efforts be used publicly again with safe reliance on the algorithms.(Almasoud & Idowu, 2024).
4. **Incorporation of more variables**: While more variables aren't always a great thing for privacy, they've proven to be beneficial when it comes to improving model performance. Spatio-temporal data for crimes has been used in an STKDE approach, which has shown itself to be better than many approaches currently being considered. However, this process is resource intensive, as we found out during our implementation of the algorithm.

Evaluation and Reflection

Impact Assessment

The societal impact of predictive policing extends beyond its technical flaws. By reinforcing existing inequalities, these systems risk deepening public mistrust in law enforcement (Purves, 2022). Additionally, the lack of community engagement in the design and deployment of predictive systems further alienates affected populations (Hung & Yen, 2020).

Proposed Solutions

1. **Community-Centric Design**: Engaging affected communities in system design can foster trust and ensure fairness (Davis et al., 2022).
2. **Regular Audits**: Conducting audits of predictive systems for bias and effectiveness is crucial for maintaining accountability (Gstrein et al., 2019).
3. **Alternative Interventions**: Emphasizing non-enforcement strategies, such as social services and community programs, can mitigate the over-reliance on predictive tools (Purves, 2022).
4. **Proper Feature Engineering & Model tuning**: Once we have the data, it is important to take only the parts which can be deemed ethical to be used and not be identifying or even quasi-identifying, and ensure that they're pre-processed properly before feeding them to a model. Once we see what the model has given us, it's important to tune it so that it has the right idea from the data. In our case, failure to tune the model properly is what caused the model to be inaccurate, since it overfit

and gave too much importance to a variable which didn’t really matter as much. Moreover, the dataset also wasn’t anonymized right, despite the classification of crimes and location, but the addition of noise did seem to help the case.

Conclusion

While predictive policing can potentially revolutionize law enforcement, its ethical and societal challenges are too big to be ignored.If we are to address this problem, we must incorporate all stakeholders and come up with approaches both technically and ethically viable. Future research can be around refining fairness metrics, innovation in approaches to improve privacy and fairness, and performance comparison for multiple approaches.

References

Almasoud, A. S., & Idowu, J. A. (2024). Algorithmic fairness in predictive policing. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00541-3

Davis, J., Purves, D., Gilbert, J., & Sturm, S. (2022). Five ethical challenges facing data-driven policing. *AI and Ethics*, 2(3), 185–198. https://doi.org/10.1007/s43681-021-00105-9

Gstrein, O. J., Bunnik, A., & Zwitter, A. J. (2019). Ethical, legal, and social challenges of predictive policing. *Cutting Crime Impact Project*. https://ssrn.com/abstract=3447158

Hung, T.-W., & Yen, C.-P. (2020). On the person-based predictive policing of AI. *Ethics and Information Technology*, 23(3), 165–176. https://doi.org/10.1007/s10676-020-09539-x

Mugari, I., & Obioha, E. E. (2021). Predictive policing and crime control in the United States and Europe. *Social Sciences*, 10(234), 1–14. https://doi.org/10.3390/socsci10060234

Purves, D. (2022). Fairness in algorithmic policing. *Journal of the American Philosophical Association*, 7(4), 741–761. https://doi.org/10.1017/apa.2021.39

Susser, D. (2021). Predictive policing and the ethics of preemption. *In B. Jones & E. Mendieta (Eds.), The Ethics of Policing: New Perspectives on Law Enforcement*.