

Conclusion Document

Group 5

December 7, 2024

1 Overview

For the deep dive project, we choose to analyze the “Chicago Crime Dataset” with the primary objective of being able to predict future types of crime. As a secondary goal, we identify that if we are able to reasonably predict the types of crime, we can then perform evaluations such as feature importance to evaluate the importance and predictive power of the different features. This feature importance evaluation can be critical as an analysis tool in identifying predictive features, informing, and supporting the need for intervention in different communities or times.

The rest of the conclusion document will briefly discuss motivations, key components in the model architecture, design, and finally some discussion on the findings.

2 Motivation

A successful and accurate predictive model has major decision making implications on the part of the police department. With an accurate predictive model, police departments can attempt to anticipate crime types at different locations. With extreme confidence in the model, this might translate to the police allocating resources in particular locations at different times of the day and year. Even without an extremely confident model, this can inform the need to provide different types of intervention such as informative interventions. That is, if the police department is aware that a particular type of crime spikes at a particular time of day or year, they can share important information that can aid in combating this crime. For example, informative posters on how to prevent burglaries, notifications of the dangers of walking home alone at a particular time of day, etc.

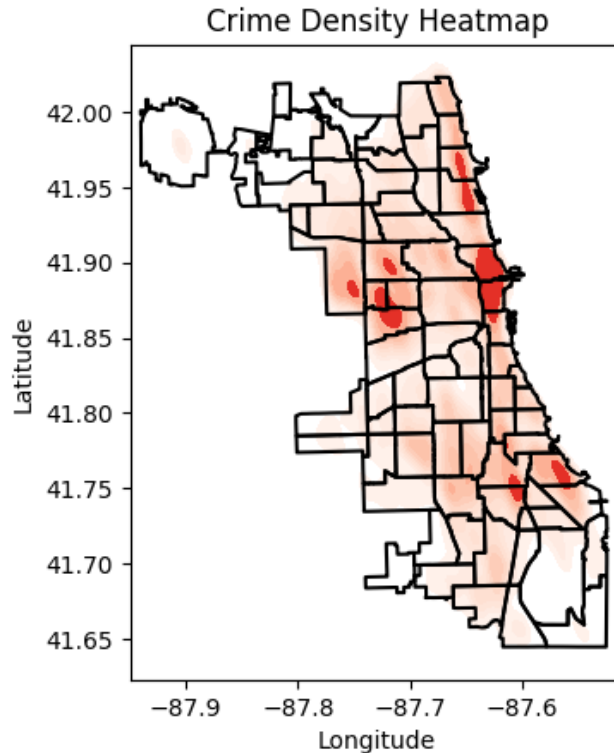
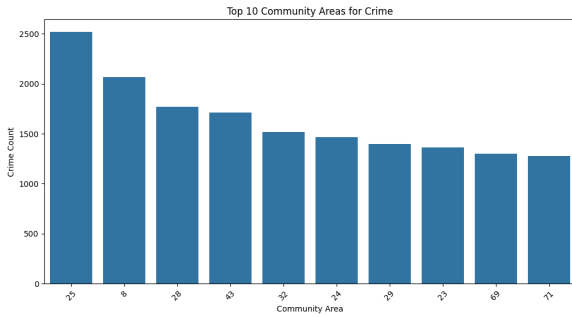
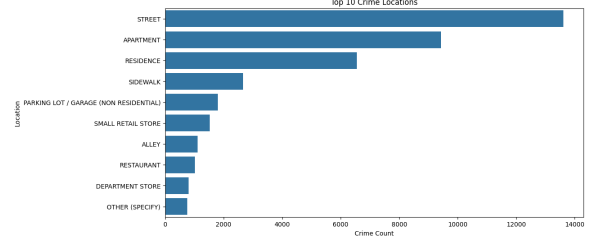


Figure 1: Crime density heat-map overlaid on Chicago divided by county.

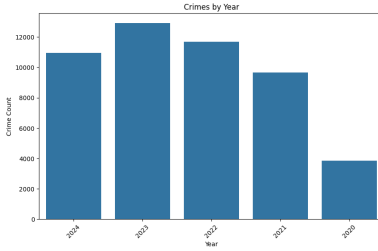


(a) Crime Count at Community Codes

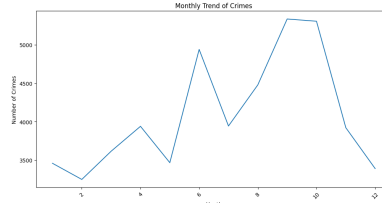


(b) Crime Count versus location

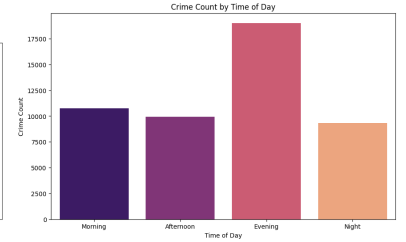
Figure 2: Spatial Analysis of Crime



(a) Crime by Year



(b) Crime by Months



(c) Crime by Time

Figure 3: Temporal Graphs of Crime

3 Data Exploration

By looking at the dataset through a data exploration, trends and commonalities arise to lead us to believe that a deep learning model will be able to learn something. There are a few aspects of the data that we explored: spatial and temporal. Within each of these, patterns and hot-spots even visible to the average person arises. In general, the analysis give the intuition that both spatial and temporal features are going to have a strong predictive power on future crime.

3.1 Spatial.

We begin our discussion by discussing and showing the crime density heat map overlaid the city and suburbs of Chicago in Fig. 1. Through this figure, we can see four major regions whose crime dominates the frequency of crime in the region. Fig. 2 shows the crime counts measured against spatial values such as the community code and the location. These graphs demonstrate that there are clear relationships between the location and the crime.

3.2 Temporal.

Fig. 3 demonstrate clear trends in the temporal information. This provides the intuition that the temporal values are going to have predictive power. In particular, the crime by the month and crime by time of day demonstrate the most predictive power. We see both that during the winter months, crime tends to fall. As the weather heats up, the crime rate increases. Crime by time of day seems to be pretty stead with the exception of the evening. Crime in the evening is by far the most common time for crime.

4 Model

In this section, we briefly discuss some key aspects of the model and technical challenges encountered.

4.1 Temporal Patterns Matter

The LSTM layer successfully records crime trends across time, demonstrating that crimes might follow predictable patterns according to the time of day or night, such as increased activity on weekends or at night.

4.2 Robustness Against Noise

The model's capacity to generalize is enhanced by dropout and batch normalization, underscoring the significance of managing unpredictability in actual crime data, such as irregularities in location reporting.

4.3 Efficiency in Decision-Making

The Adam optimizer guarantees faster learning, allowing for quicker insights into crime categories. This is essential for prompt actions in urban safety planning.

4.4 Feature Transformations are Key

By embedding category characteristics like "crime type" and normalizing geographical coordinates, the model is able to identify subtle contextual and spatial correlations, including recurring hotspots for particular crimes.

4.5 Scalability for Urban Analysis

The model's modular architecture allows it to be tailored to various cities or bigger datasets, opening the door for more extensive uses in predictive policing and urban safety.

4.6 Categorical Prediction

One challenge that we did not expect from the beginning was the categorical nature of crime type. As we have seen in class, performing prediction on categorical data is not simple. Even in the regression case, we see that we introduce an entirely new formatting of data, loss function, and analysis. In the deep learning setting, we encountered that the models that we discussed in class would not work out of the box and would require modifications. Albeit, there are resources for performing categorical predictions for the most popular deep models, but this required further research than we expected.

5 Results

5.1 Feature Importance

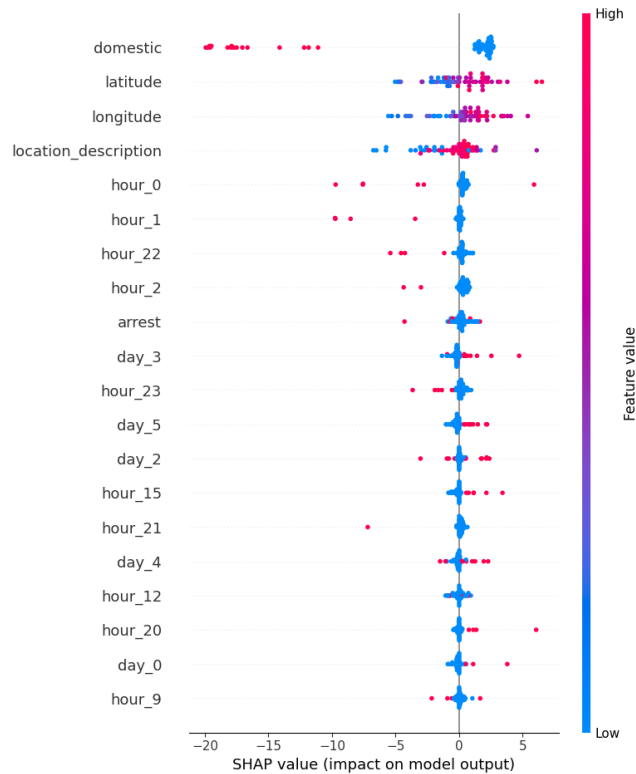


Figure 4: Feature Importance Plot Ordered by Shapley Value.

As we can see through Shapley feature importance in Fig. 4, the domestic keyword is a very strong indicator of crime type. This is likely due to the specific types of crime that can occur in a domestic context, such as assault and battery. However, this keyword’s usefulness mainly pertains to a specific subset of crime, so we need to consider other features as well. Location-related features were the next most important, which also makes sense. Certain types of crime, like robbery, are likelier in the hotspots like Jeweler’s Row, as well as in locations described as some type of store. As hours were one-hot encoded, they are treated as separate features. This allows us to see which hours were able to provide more predictive power, and we realize a trend that hours around midnight (that is, late nights and early mornings) are more important. This is likely due to increased crime data points at these hours and that certain more serious crimes are likelier at night.

5.2 Performance

After selection of optimal hyperparameters, we saw a roughly 33.25 percent accuracy rate with our deep learning model. This is a substantial accuracy, considering the noisy nature of crime data and the various classes we are trying to predict. As a baseline comparison, a logistic regression model with equally weighted classes provided an accuracy of only 2-3 percent (the unweighted model predicted theft or battery in all cases, leading to inflated accuracy of 25 percent on the test set). The Shapley plot also underlines the sophisticated use of inputs and subsequent expressivity of the model.