Enhancing Predictive Policing Through Spatio-temporal Crime Forecasting

Alice Kamau

December 2023

1 Introduction

The project aims to develop models that enhance predictive policing through spatiotemporal crime forecasting. Crime is a global problem affecting the quality of life and economic growth[4]. It's more prevalent in urban areas due to population density and economic disparities, often linked to unemployment and income. Crime analysis historically involves behavioral, psychological, and sociological aspects with established theories like 'rational choice theory', 'journey to crime', and 'routine activity theory'[3]. The growth of data in the last decade offers new opportunities for crime analysis. Data-driven approaches in law enforcement have gained popularity, aided by machine learning algorithms. These technologies identify crime patterns, influence factors, and support predictive policing efforts, saving time for law enforcement[7]. Thus, this study aims at achieving the following objectives: i. To determine the crime hotspot areas and seasons in Portland, Oregon. ii. To come up with a model to predict crime occurrences in Portland, Oregon.

2 Related Work

Different crime analysis studies have been conducted in different regions and focus on diverse aspects. A study by Wheeler and Steenbeek[7] mapped the risk terrain for crime using machine learning. Their study illustrated how long-term predictions of crime can be achieved through machine learning algorithms and random forests, focusing on robberies in Dallas, Texas. The study found out that factors used to predict crimes were non-linear and varied over space. The current study aims to come up with a model(s) to forecast crimes in Portland, Oregon. Another study by Vivek and Prathap[6] aimed to provide a spatio-temporal view of crime in India using statistical and machine learning models. They performed Time series forecasting of crime tweets by comparing the accuracy of Long Short-Term Memory (LSTM), Auto-Regressive Integrated Moving Average (ARIMA), and Seasonal Auto-Regressivee Integrated Moving Average

(SARIMA) models to determine the best model. The current study will employ some of the models used by Vivek and Prathap[6] and determine their effectiveness. Alazawi, Jiang, and Messner[5] focused on Columbus, Ohio, using residential burglary data from 1994 to 2002. Findings revealed that residential burglary in Columbus clustered at a characteristic scale of 2.2 km. Deckard and Schnell[2] examined temporal patterns between months to explore the impact of the modifiable temporal unit problem (MTUP) in St. Louis, Missouri. There was much variability in which locations were hot spots between months. Most locations were found to have complete instability or unpredictability in the months when crime does occur. In Jersey City, Caplan et al.[1] examined temporal variations in the spatial influence of environmental features on criminal behavior across microlevel places. Environmental risk factors and their spatial influences were identified for street robberies. The results suggested that there was a temporality to robbery due to the interaction between physical vulnerabilities from the built environment and the social behaviors of people at these places.

3 Methods

This project aims to come up with models that can help predict future crime occurrences in the area under study.

3.1 Sources of data

The study used datasets from the National Institute of Justice's (NIJ) 'Real-Time Crime Forecasting Challenge' for the years 2013 to 2016. Since the data is from different datasets, data integration will be done to combine all the datasets from 2013 to 2016. Data was sorted based on their timeline so that to be in a time series. Data resampling from daily to monthly intervals and feature extraction was conducted. This dataset served several purposes: a. Classifying data based on areas (x and y coordinates), crime type, and year of occurrence. b. Analyzing the crime patterns and their correlation. c. Predicting the crime hotspots with temporal crime data.

3.2 Data preprocessing

In preprocessing, duplicate records are removed, and missing places filled. Outliers will be detected and treated accordingly. Cleaning involved standardizing data formats, handling null values, and addressing any anomalies that may affect the quality of the analysis. time-series graphs are employed to represent spatiotemporal crime patterns. This aids in better understanding and interpretation of the data. Time-series forecasting techniques, specifically ARIMA (AutoRegressive Integrated Moving Average), are employed to predict future crime occurrences. This involves analyzing historical crime data to identify patterns and trends that can be extrapolated into the future. The forecasted results provide valuable insights for proactive law enforcement strategies.

3.3 Implementation plan

For descriptive statistics, basic measures such as mean, mode, median, and standard deviation were computed. The general frequency distributions for categories such as call groups, category were determined. Pandas and Matplotlib libraries were utilized to compute these statistics for numerical and frequency distributions for categorical columns. Visualizations such as bar charts, line graphs, histograms, and pie charts were employed to explore the distribution of categorical data. Heat maps and scatter plots were utilized to visualize the spatial distribution of incidents based on x and y coordinates. Matplotlib was used for analytical visualization, while libraries such as Seaborn and Folium were applied for geographic data visualization. Geospatial data mining techniques were used to identify spatial patterns and clusters in the dataset. Clustering algorithms like K-means were employed to identify areas with high incident densities. Geopandas, Scikit-learn, and Matplotlib libraries were considered in this step. The study looked for trends, seasonality, and anomalies of data grouped and plotted over time. Forecasting involved a comparative analysis of ARIMA models to determine the most accurate forecasting model for crime. Libraries such as Pandas and Statsmodels were applied in this phase.

4 Results and Discussion

4.1 Distribution of crime by category

In Figure 1, the bar graph illustrates the crime count categorized into three main types: street crimes, motor vehicle theft, and burglary, with all other types combined under "others." The data indicates that "others" have the highest total crimes, suggesting a diverse range of criminal activities beyond the three main categories. The prominence of the "others" category implies a significant diversity in criminal activities not limited to the primary crime types identified, namely street crimes, motor vehicle theft, and burglary. This calls for a more nuanced approach to law enforcement, considering the wide array of criminal behaviors falling under the "others" category. Allocating resources and developing strategies to address this diverse range of crimes will be essential for an effective law enforcement response.

4.2 Distribution of call groups

Figure 2 provides insight into the distribution of crime calls based on different call groups. Disorder, non-criminal, and property crime are the prominent call groups, with disorder having the highest count. Conversely, person crime has the lowest count among the call groups. The high count in the disorder category highlights a potential area of concern that demands attention from law enforcement. Understanding the distribution of calls across various groups is crucial for

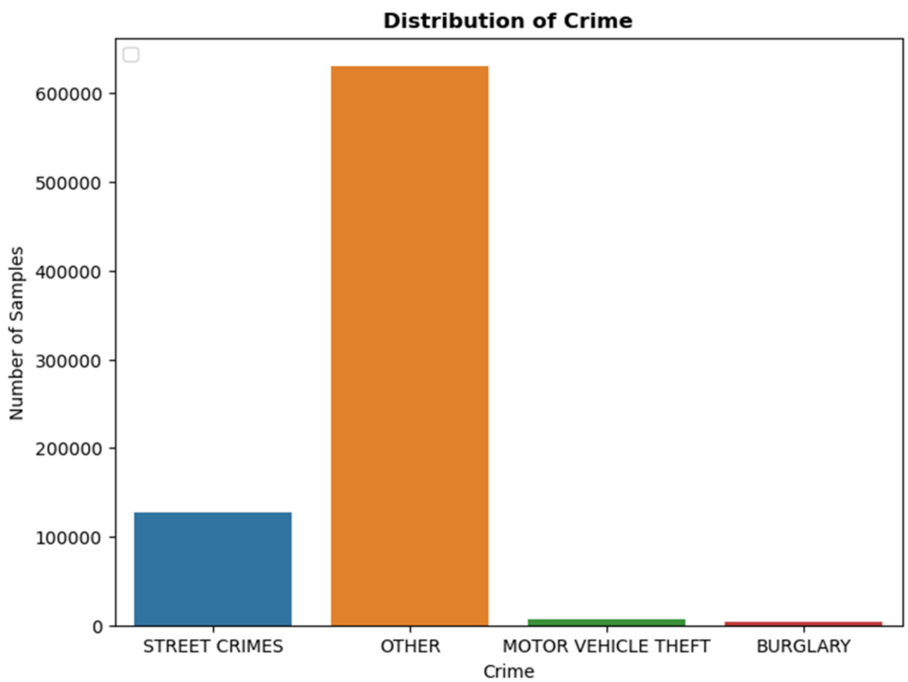

Figure 1: Category of Crimes

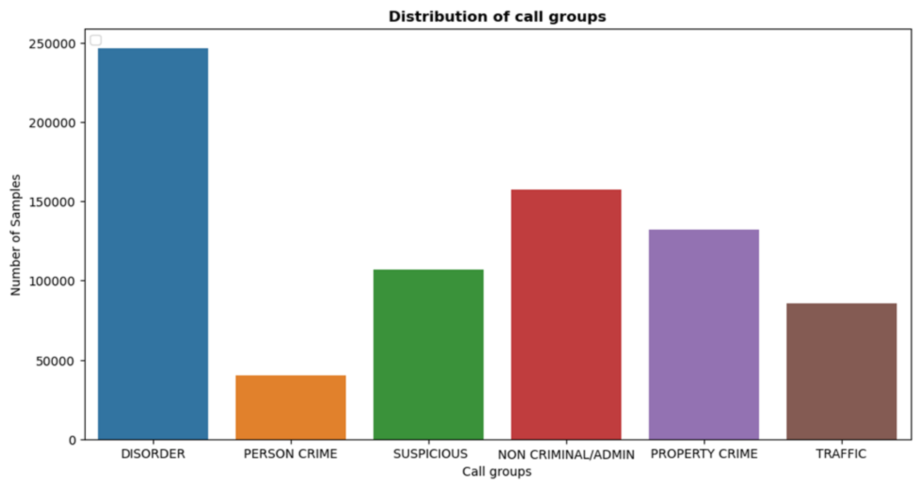

Figure 2: Call groups

prioritizing response efforts. While disorder, non-criminal, and property crime are predominant, the lower count in person crime suggests a less frequent but potentially more severe category, necessitating focused attention.

4.3 Time Series Analysis

Figures 3 and 4 present the time series analysis of street crimes over a four-year period, both on a daily and monthly basis. The trend indicates a seasonal pattern, with crime peaking during the summer months and fluctuating during winter. This information can be crucial for law enforcement to allocate resources effectively, considering the temporal variations in crime rates. Seasonal adjustments to resource allocation and strategies may be necessary to effectively address the changing dynamics of street crimes throughout the year.

4.4 Time series forecasting

Figure 5 displays the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots used for identifying the order of the ARIMA model. The chosen model's order is (3, 2, 11). These plots help assess the correlation between time series data and its lags, aiding in the selection of appropriate parameters for the forecasting model.

4.5 ARIMA Model

The Root Mean Squared Error (RMSE) of 625430.89, as shown in Figure 6, indicates the performance of the ARIMA model. The high RMSE suggests that the model may not be the best fit for the data, and alternative models

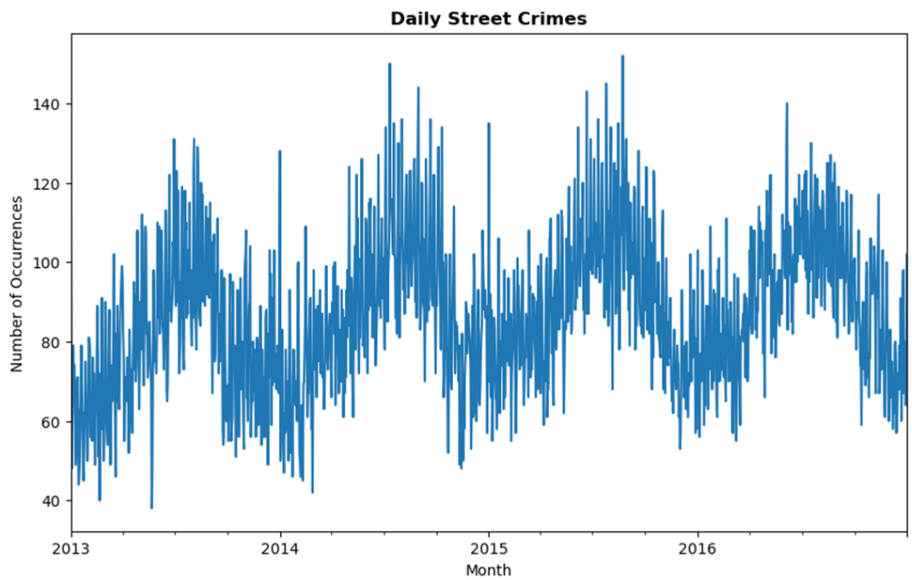

Figure 3: Daily street crimes

Figure 4: Monthly street crimes

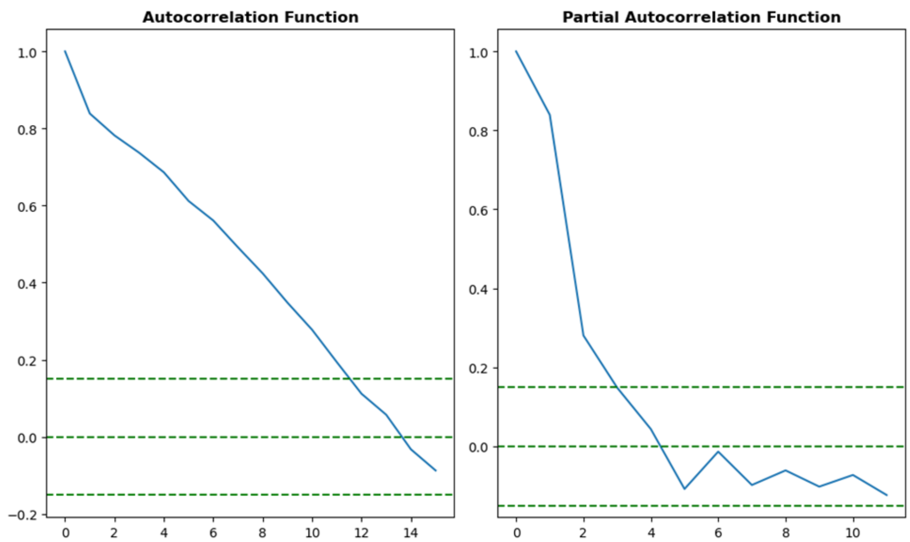

Figure 5: ACF and PACF

like SARIMA or LSTM could be explored. Limited time constraints may have impacted the thorough exploration of alternative models.

4.6 Multi-step forecasting

The ARIMA model achieved a Mean Absolute Percentage Error (MAPE) of 8.44 percent in predicting street crimes. This metric provides a measure of the accuracy of the forecasting model, with a lower MAPE indicating better predictive performance. The result suggests that the ARIMA model is relatively accurate in forecasting street crimes based on the available data. While ARIMA model provides a reliable estimate, the discussion throughout emphasizes the importance of considering alternative models to potentially enhance predictive capabilities.

5 Conclusion

In conclusion, the presented visualizations and analyses offer a multifaceted perspective on the dynamics of crime, providing essential insights for law enforcement and policymakers. The bar graphs highlight the diversity of criminal activities, emphasizing the significance of addressing not only major crime categories but also the nuanced behaviors falling under the "others" category. The distribution of call groups underscores the prevalence of disorder-related incidents, guiding law enforcement in prioritizing response efforts. The time series analysis reveals temporal patterns in street crimes, with distinct seasonal varia-

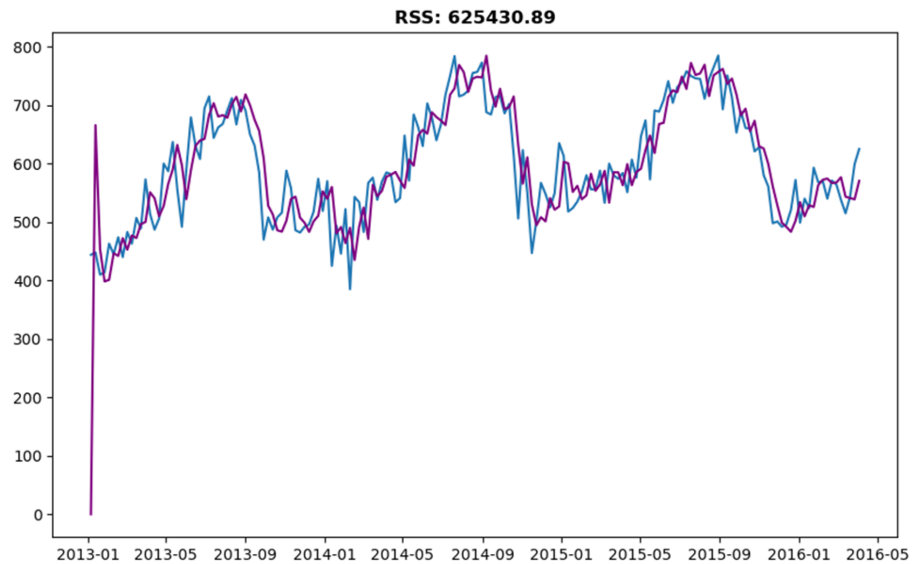

Figure 6: ARIMA Model

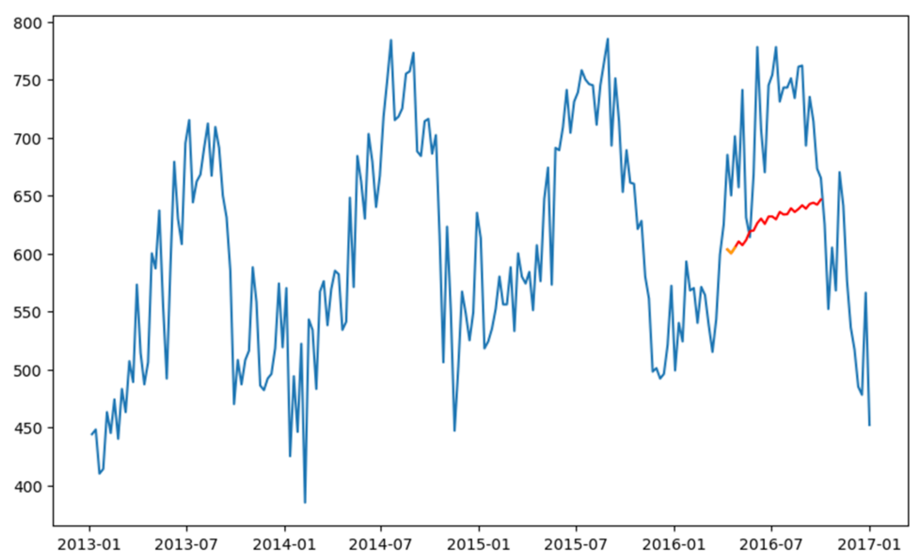

Figure 7: Predictive model

tions that necessitate adaptive law enforcement strategies. Despite the practical utility of the ARIMA model in forecasting, the high Root Mean Squared Error (RMSE) prompts consideration of alternative models like SARIMA or LSTM for improved accuracy. The decision to stick with ARIMA due to time constraints emphasizes the challenges of real-world applications where resource limitations may impact the thorough exploration of diverse forecasting methodologies. While the Mean Absolute Percentage Error (MAPE) indicates reasonable accuracy in street crime prediction using the ARIMA model, the discussion consistently underscores the need for ongoing research and experimentation with different models. Future endeavors should focus on refining forecasting techniques to better align with the dynamic nature of criminal activities, ensuring law enforcement agencies can proactively and effectively respond to emerging trends in crime.

References

[1] Joel M. Caplan, Christine H. Neudecker, Leslie W. Kennedy, Jeremy D. Barnum, and Grant Drawve. Tracking risk for crime throughout the day: An examination of jersey city robberies. *Criminal Justice Review*, 46(2):259–273, 2021.

[2] Mica Deckard and Cory Schnell. The temporal (in)stability of violent crime hot spots between months and the modifiable temporal unit problem. *Crime & Delinquency*, 69(6-7):1312–1335, 2023.

[3] Frank E. Hagan and Leah E. Daigle. *Introduction to Criminology: Theories, Methods, and Criminal Behavior*. SAGE Publications, 2019.

[4] Anja P. Jakobi. *Crime, Security and Global Politics: An Introduction to Global Crime Governance*. Macmillan International Higher Education, Red Globe Press, 2020.

[5] Steven F. Messner Mohammed A. Alazawi, Shiguo Jiang. Identifying a spatial scale for the analysis of residential burglary: An empirical framework based on point pattern analysis. *PLoS ONE*, 2022.

[6] Meghashyam Vivek and Prathap Rudra Boppuru. Spatio-temporal crime analysis and forecasting on twitter data using machine learning algorithms. *SN Computer Science*, 4, 05 2023.

[7] Andrew Wheeler and Wouter Steenbeek. Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 37, 06 2021.