Crime Rate Forecasting in Chicago: A Time Series Analysis

Authors: Asad Adnan, Muhammad Ahmad, Brian Murphy

Date: February 17, 2025

1. Introduction & Objective

Chicago has long grappled with persistent crime, particularly in its South and West sides, where theft and battery dominate. These crimes erode community safety, hinder economic growth, and strain law enforcement resources. Accurate crime forecasting can transform this challenge into an opportunity by empowering law enforcement and policymakers to anticipate trends, allocate resources efficiently, and implement proactive crime prevention strategies.

This study leverages historical crime data to predict monthly arrest counts for theft and battery—Chicago's most prevalent offenses—using advanced time-series forecasting models. Our objective is to deliver reliable forecasts that inform proactive strategies, from patrol deployment to community interventions, ultimately enhancing public safety. This study underscores the need to integrate data-driven decision-making in public administration to ensure that citizens receive the best possible protection and support.

2. Data Description

Source:

The dataset is sourced from the Chicago Data Portal's "Crimes - 2001 to Present" (https://catalog.data.gov/dataset/crimes-2001-to-present). We primarily used data from 2001 to December 2023, encompassing over 7 million records.

Processing:

- We filtered for theft (1.75M cases) and battery (1.51M cases) arrests, which are among Chicago's most prevalent crimes.
- Data was aggregated into monthly arrest counts to identify temporal trends.
- Geospatial elements (Block, Latitude) were excluded for a purely time-series approach. However, future work could integrate spatial analysis for more granular forecasting.

Rationale:

Theft and battery are key indicators of public safety concerns in Chicago. Unlike some other major cities like New York, where aggressive policy changes have led to reduced crime rates, Chicago continues to struggle with fluctuating crime patterns. By applying time-series analysis, we can extract insights from historical data and generate reliable predictions to support law enforcement agencies, policymakers, and urban planners in targeting crime strategically and deploying resources efficiently.

---

3. Research Questions

1. How have monthly theft and battery arrests trended in Chicago from 2001 to 2023?
2. Which time series model delivers the most accurate crime forecasts?
3. How can these forecasts optimize law enforcement strategies and resource allocation?
4. Who benefits from these insights (police, policymakers, residents), and how can they be applied?
5. What policy actions arise from our findings?

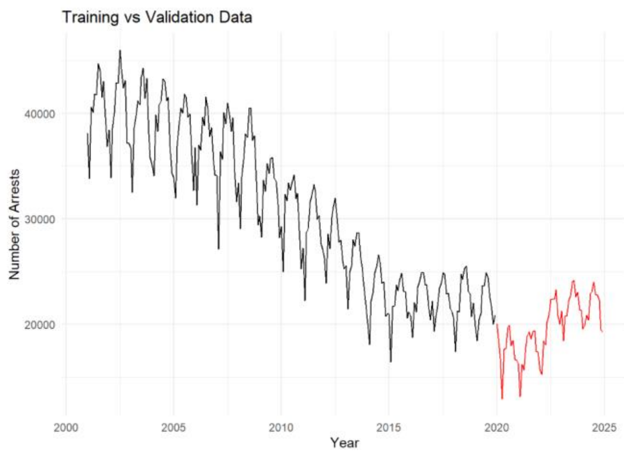
---

4. Data Visualization

A time-series plot of monthly arrest counts (2001–2023) reveals a ~50% decline over time, with notable seasonal peaks in warmer months (June–August) and significant dips during harsh winters. Additionally, a sharp drop in crime during COVID-era lockdowns (2020–2021) is evident.

Additional Insights:

- Crime rates remain cyclical, suggesting predictable seasonal fluctuations that can be leveraged for strategic planning.
- External factors such as unemployment rates, economic downturns, and major city events (e.g., protests, sporting events) impact crime trends.
- Demographic shifts and urban development influence long-term crime trends.

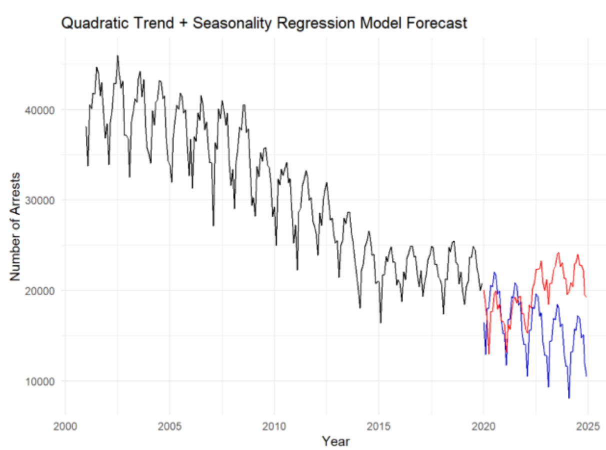


5. Forecasting Models & Methodology

We partitioned the dataset into training (2001–2019) and validation (2020–2023) sets and tested five forecasting models along with three ensemble approaches. The accuracy of each model was measured using Mean Absolute Percentage Error (MAPE).

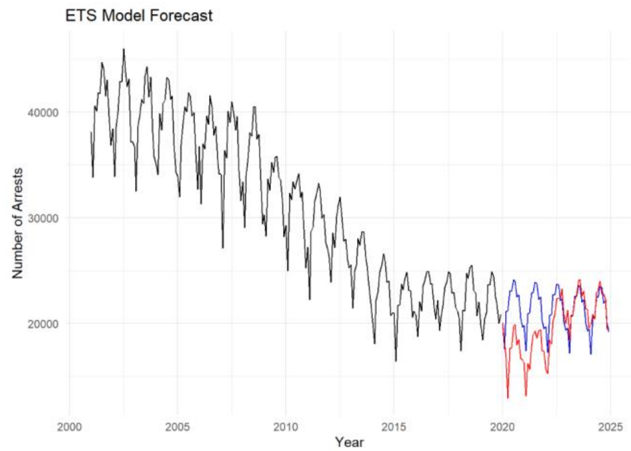
5.1 Regression Model (Quadratic Trend + Seasonality)

Method: tslm(train.ts ~ trend + I(trend^2) + season)
MAPE: 22.27%
Insight: This model captures long-term trends with polynomial terms, making it effective in detecting general patterns but ineffective in responding to abrupt crime rate fluctuations. The high MAPE indicates that its assumptions about crime trends lead to significant forecasting errors, particularly during the pandemic and post-pandemic shifts.



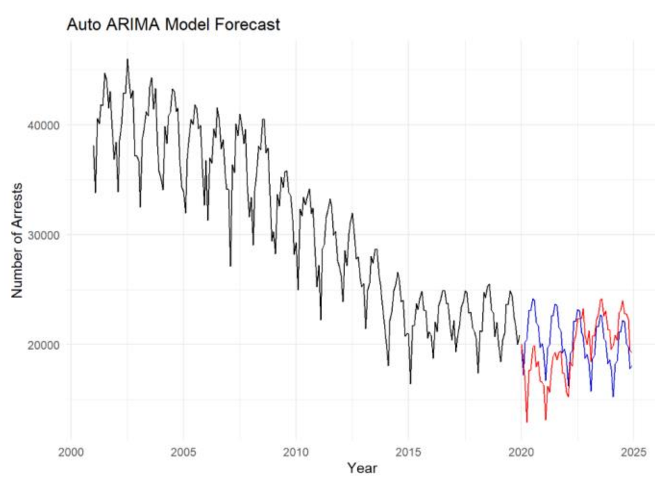
5.2 Exponential Smoothing (ETS)
Method: ets(train.ts, model = "ZZZ")
MAPE: 12.59%

Insight: ETS performed well in forecasting short-term fluctuations by weighting recent data more heavily. The seasonal adjustment aspect improved accuracy, especially in predicting summer spikes. However, it still struggled with irregular shocks like law enforcement policy shifts and unforeseen socio-political events.



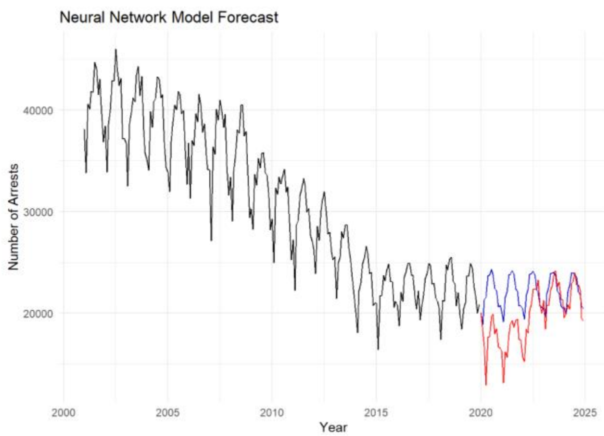
5.3 Auto ARIMA

Method: auto.arima(train.ts)
MAPE: 13.85%
Insight: Auto ARIMA efficiently captured autoregressive and moving average components, making it strong in medium-term forecasting. It effectively accounted for cyclical trends but struggled in cases where crime rates changed due to external shocks, such as the pandemic or sudden economic downturns.



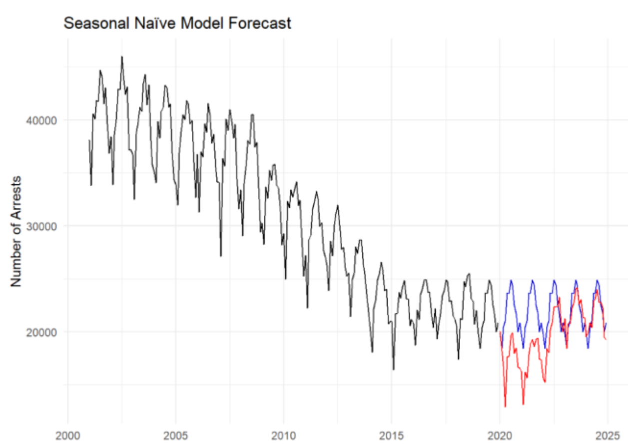
5.4 Neural Network (NNAR)

Method: nnetar(train.ts, repeats = 20, p = 11, P = 1, size = 7)
MAPE: 14.83%
Insight: NNAR provided insights into non-linear patterns and was useful for identifying crime trends that didn't fit traditional statistical models. However, its predictive accuracy remained lower than ETS and ARIMA due to the complexity of fine-tuning hyperparameters and its sensitivity to outliers.



5.5 Seasonal Naïve Model

Method: snaive(train.ts, h = nValid, level = 0)
MAPE: 13.74%
Insight: The Seasonal Naïve model's performance reaffirmed the strong seasonal component of crime rates. While it served as a strong benchmark, its inability to account for long-term shifts limited its predictive power beyond simple cyclical forecasts.



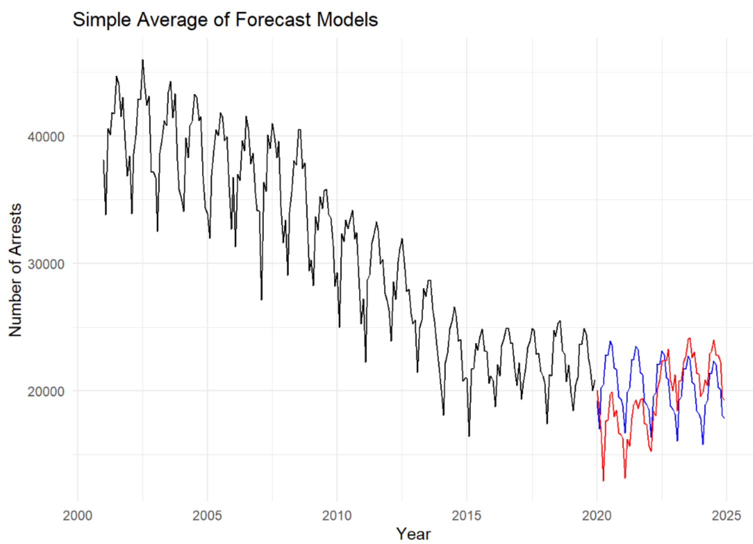
5.6 Forecast Combinations

We combined forecasts from the five models using the following three approaches.

Simple Average:

MAPE: 13.23% (MAE: 2441 arrests)

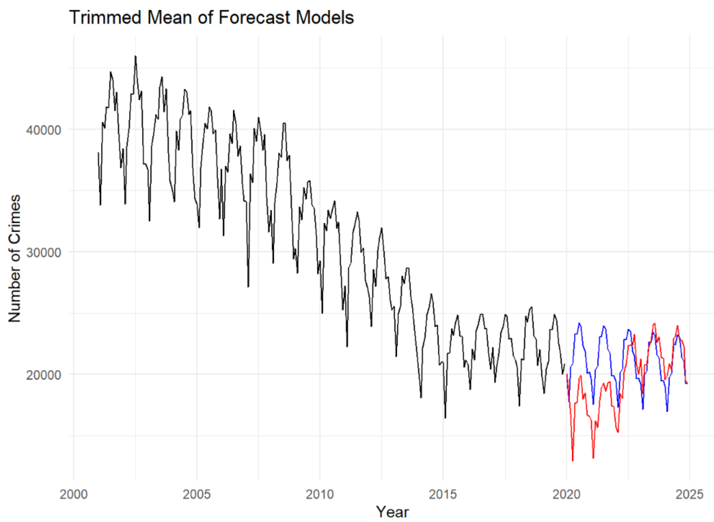
- This method takes the mean of all five model outputs to generate a single prediction. While it provides a balanced estimate, it does not account for variations in model performance. As a result, it is less effective when individual models have vastly different levels of accuracy.



Trimmed Mean:

MAPE: 12.79% (MAE: 2285 arrests)

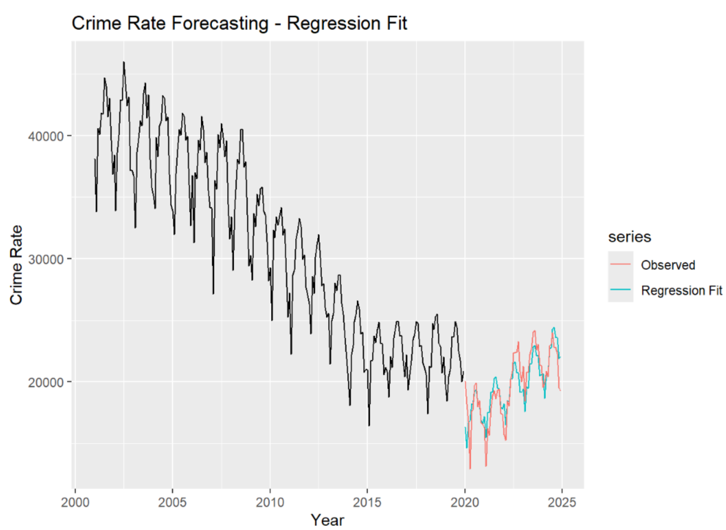
- By excluding extreme high and low forecasts, this method reduces the impact of outliers and enhances stability. It improves accuracy by mitigating the effect of erratic predictions but still lacks an adaptive mechanism to weigh stronger models more effectively.



Regression-Based Combination:

MAPE: 5.72% (MAE: 1071 arrests)

- This approach assigns different weights to each forecasting model based on their past performance, leading to significantly better accuracy. By leveraging the strengths of each model, this method adapts dynamically to changing trends, making it the most effective forecasting technique among those tested.



Limitations: Arrest-only focus may underrepresent total crime; excluding external factors (e.g., socioeconomic shifts, weather) limits depth. Future models could incorporate these.

6. Policy Implications & Recommendations

Strategic Applications:

- Peak Patrols: Increase police presence by 25% in June–August in high-crime zones based on seasonal forecasts.
- Resource Planning: Use ETS or ARIMA for quarterly budgeting of community policing and intervention programs.
- Dynamic Adjustments: Update crime strategies monthly using Regression Combination forecasts to adapt to real-time trends.
- Community Engagement: Collaborate with local organizations, youth programs, and neighbourhoods watch groups to address root causes of crime. By involving community leaders, law enforcement can build trust and encourage proactive reporting of suspicious activity.
- Long-Term Strategy: Invest in predictive analytics tools and expand the dataset to incorporate socioeconomic factors, weather conditions, and emerging crime trends for continuous improvement in crime prevention efforts.

7. Conclusion

Crime forecasting is not just a theoretical exercise—it has real-world implications that can enhance public safety, improve law enforcement efficiency, and support policy development. Our study demonstrates that time-series forecasting models, particularly regression-based combinations, can significantly improve crime prediction accuracy. The regression-based combination model (MAPE: 5.72%) was the most reliable, effectively balancing the strengths of all other methods while minimizing weaknesses.

From a practical perspective, these forecasts allow law enforcement agencies to allocate resources proactively rather than reactively. Policymakers can use these insights to develop data-driven crime prevention programs, optimizing budget allocations and increasing the effectiveness of community outreach initiatives. Additionally, the findings emphasize the role of seasonality and external factors in shaping crime trends, reinforcing the need for adaptive crime prevention strategies.

Moreover, community involvement plays a crucial role in crime reduction. Strengthening partnerships between law enforcement, local organizations, and residents can help implement crime prevention programs tailored to specific neighbourhoods. Forecasting models should not only serve as decision-making tools for law enforcement but also foster a collaborative approach to crime reduction that includes social initiatives and public engagement.

Future work can further refine these models by incorporating socioeconomic indicators, spatial analysis, and deep learning techniques to improve predictive performance. By integrating statistical precision with real-world applications, we present a data-driven framework for a safer Chicago, ensuring that crime prevention strategies are not just reactive but proactively informed by predictive analytics.

---

8. References

- Chicago Data Portal. (2024). *Crimes - 2001 to Present.* Retrieved from https://catalog.data.gov/dataset/crimes-2001-to-present
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice.* OTexts.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control.* Wiley.
- Chatfield, C. (2004). *The Analysis of Time Series: An Introduction.* CRC Press.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting.* Wiley.
- R Documentation. (2025). *Forecasting functions in R.* Retrieved from https://www.rdocumentation.org

9. Appendix

9.1 Model Implementation in R

Below is an expanded view of some of the important bits of our R implementation for forecasting crime trends:

Loading and Preprocessing Data

```
library(forecast)
library(tseries)
library(ggplot2)

# Load crime data
crime_data <- read.csv("Chicago_Crime_Data.csv")

# Convert to time series
time_series <- ts(crime_data$Arrests, start = c(2001,1), frequency = 12)

# Check stationarity
adf.test(time_series)
```

Building Forecasting Models

```
# Exponential Smoothing Model
ets_model <- ets(time_series)
forecast_ets <- forecast(ets_model, h=24)

# ARIMA Model
auto_arima_model <- auto.arima(time_series)
forecast_arima <- forecast(auto_arima_model, h=24)

# Seasonal Naïve Model
snaive_model <- snaive(time_series, h=24)
forecast_snaive <- forecast(snaive_model)
```

Evaluating Model Performance

```
# Compare model accuracy
ets_accuracy <- accuracy(forecast_ets, time_series)
arima_accuracy <- accuracy(forecast_arima, time_series)
snaive_accuracy <- accuracy(forecast_snaive, time_series)

# Display accuracy results
```

```
print(ets_accuracy)
print(arima_accuracy)
print(snaive_accuracy)
```

Visualization of Forecasts

```
# Plot Forecasts
autoplot(time_series) +
  autolayer(forecast_ets, series="ETS", PI=FALSE) +
  autolayer(forecast_arima, series="ARIMA", PI=FALSE) +
  autolayer(forecast_snaive, series="SNaive", PI=FALSE) +
  ggtitle("Crime Forecasts using Different Models") +
  ylab("Number of Arrests") + xlab("Year")
```

For the full implementation and interactive visualizations, please refer to the accompanying **FP-TSF.html** file.