Crime Rate Forecasting using Time-Series Models: A Comparative Study of SARIMA and Prophet

Samah Samir
Computer Science, Faculty of Computers and Information, Menoufia University, Egypt

Hamdy M. Mousa
Computer Science, Faculty of Computers and Information, Menoufia University, Egypt

Eman M. Mohamed
Computer Science, Faculty of Computers and Information, Menoufia University, Egypt

ABSTRACT
Crime is a major problem in all societies, impacting the economic and social lives of individuals and communities. Therefore, it is important to study and analyze the influencing factors and the various relationships between the motives for different crimes to prevent their recurrence in the future. Crime prediction is a method of studying the causes and motives of crime and predicting the times and places of its occurrence to reduce the occurrence of expected crimes in the future. This study aims to analyze the Los Angeles dataset and investigate the impact of certain factors on the crime rate. We used Time-series forecasting to predict future crime trends. Among Time-series forecasting models, we used SARIMA (Seasonal Auto Regressive Integrated Moving Average) and Prophet model to predict the future crime rate. The study succeeded in extracting some results by analyzing a set of factors (social and economic) that affect crime rates. The study also succeeded in training the model to predict future crimes for the years 2026 and 2027 by using historical data from 2020 to February 2025 to train the model.

Keywords
Crime prediction; Crime data analysis; time-series forecasting; ARIMA; SARIMA; Prophet.

1. INTRODUCTION
Crime is a negative phenomenon that occurs worldwide in both developed and underdeveloped countries. Criminal activities can negatively impact a country's overall economy and affect the quality of life and daily activities of its population, leading to significant social and economic problems. Crime and criminal activity can negatively impact both the public and private sectors, placing additional burdens on them and affecting society as a whole. Different types of crime may be associated with different factors. In general, crimes are caused by various conditions, including the social and economic environment, human nature, poverty, high population density, unemployment, gender inequality, and illiteracy. All of these factors can contribute to increased violent crime. Underdeveloped and overpopulated cities are closely associated with high crime rates [1]. Community stability and security depend on reducing crime so that people can live in peace, while communities that suffer from widespread crime cannot progress socially and economically. The overall safety of communities and individuals is an important factor in providing safe environments when people travel, move to new locations, or engage in new economic activities. Therefore, analyzing crime reports and statistics is essential to improving community and individual safety. Solving crimes is a challenging task that requires human intelligence, expertise,

and time. Data analysis is one of the most effective techniques for solving crimes and predicting future incidents. Crime prediction has gained significant importance in recent years, helping investigative authorities tackle crime using computer-based [13]. Therefore, better predictive algorithms are needed to guide authorities toward criminals. Numerous studies have been conducted to predict crime types and rates, identify criminals' profiles, age groups, and hotspots using historical crime datasets from various regions [14].

Analyzing historical crime data is one of the most effective techniques for predicting when and where crimes will occur in the future [2]. Forecasting future crimes is the process of identifying changes in crime rates from year to year and anticipating these changes in the future. Crime can also be predicted by studying historical crime data and identifying specific factors that influence criminal activity; numerous studies in this field have been conducted.

This study aims to help police stations and relevant authorities predict future crimes by applying several techniques to real datasets for Los Angeles. This dataset reflects incidents of crime in the City of Los Angeles from 2020 to February 2025.

This study was divided into four stages, and each stage will be explained in detail. The four stages are: 1- Cleaning and preparing the crime dataset for analysis.2- Analyze crime dataset depending on some factors such as trends over time (by year, by month, by day), by region, victim age, and Victim sex.3- Study the impact of economic and social factors and major events on crime rates.4- Using time-series forecasting to predict future crime trends.

The results of this study show that months ('January', 'February', 'March', 'April') in 2024 are the highest crime counts. The highest occurrence of crime count (255106 crimes) is in the year 2024. The most common crime type is vehicle stolen. The region of Central has the highest occurrences of crimes. The highest rate of committing crimes was among victims aged 20 to 40, and the highest rate of committing crimes was among male victims.

The model predicted the crime rate for the years 2026 and 2027. The remainder of the research is organized as follows. The second section will discuss, analyze, and compare relevant work. The third section presents the proposed methodology and data sets. The fourth section will analyse the data set. The fifth section will discuss the results. Finally, the sixth section presents the conclusion.

2. RELATED WORK
In the current literature, researchers have focused their attention on analyzing criminal activities primarily from the perspectives of place and time. In [3], the paper aimed to predict time series using a deep learning model. Neural Basis Expansion Analysis for Time Series - Recurrent Neural Networks (N-BEATS-RNNs) are proven ensemble models for time-series prediction. Future trends were predicted more accurately by using the NBeats algorithm to build an effective model. The paper was applied to the Sacramento crime dataset. This study stands out from other

studies in the accuracy of the data used, as it is real data. The RNN-LSTM model applied to their univariate dataset demonstrated better performance based on the MAE measure compared to other work in the field of time-series forecasting.

In [4], They used data from two different cities: Chicago and Los Angeles. A variety of machine learning algorithms were utilized, including logistic regression, multilayer perceptron (MLP), naive Bayes, k-nearest neighbors (KNN), extreme gradient boosting (XGBoost), support vector machine (SVM), decision tree, and random forest. Additionally, time-series analysis was conducted using autoregressive integrated moving average (ARIMA) models and long short-term memory (LSTM) networks. The study shows that LSTMs performed adequately in time-series analysis in terms of root mean square error (RMSE) and mean absolute error (MAE) across both datasets. Exploratory data analysis identifies more than 35 types of crimes and indicates a year-on-year decrease in crime rates in Chicago and a slight increase in crime rates in Los Angeles. It was discovered that the lowest number of crimes occurred in February compared to other months. It was also shown that the overall crime rate in Chicago will continue to rise moderately in the future, with the potential for a decline in the coming years. It was also shown that the crime rate in Los Angeles has increased significantly, based on the ARIMA model. Among various algorithms, KNN exhibits the highest accuracy in Los Angeles, while XGBoost achieves notable accuracy on the Chicago dataset. In [5], the researchers based their study on the analysis of a variety of variables, including location, time, types of crimes committed, and demographic variations, using machine learning algorithms, statistical methods, and data mining. The primary goal was to provide law enforcement with useful information to enable them to identify crimes and maintain the public safety of countries and individuals. Data was collected from various sources, including crime reports, socioeconomic factors, arrest records, and environmental factors such as urbanization and weather trends. Random forests were used to understand how crime trends have changed over time. The study successfully created a predictive model that can predict the probability of arrests based on various variables, based on historical crime data. A random forest classifier was used to ensure robust model performance. Matplotlib and other visualization tools were used to provide a clear picture of the model's performance. In [6], the authors presented a solution to predicting future crimes using machine learning algorithms to classify a large dataset from the city of Chicago. A set of algorithms, including Random Forest, SVM, K-NN, and MLP, was used to solve this problem. The primary goal of this study was to find the best algorithm in terms of accuracy and time to solve the problem. After applying the proposed model to several machine learning algorithms, the study concluded that the best algorithm for predicting crimes, in terms of accuracy, was the Support Vector Machine (SVM) algorithm, which achieved an accuracy of 0.99997 in 1100 seconds. This was followed by the K-Nearest Neighbors (K-NN) algorithm, with an accuracy of 0.999976 and a processing time of 2450 seconds. The Random Forest algorithm came next, with an accuracy of 0.999996 and a time of 1420 seconds, while the Multilayer Perceptron (MLP) algorithm had an accuracy of 0.995886 and a processing time of 1346 seconds. The SVM algorithm is the best algorithm for achieving the highest accuracy in the shortest time. In [7], the author utilized criminal activity data from Bole Sub-City, Addis Ababa, Ethiopia. The primary objective of this study was to identify and examine the relationship between time, crime types, and locations using a neural network model for time-series data, the Long Short-Term Memory (LSTM) network. The research was divided into five phases: the first was data collection. The second was data processing to facilitate its use. The third phase was used to predict crime incidents. The fourth phase was the dataset cleaning and transformation. The fifth phase was the extraction of the model for further analysis. Experiments have shown that the LSTM model provides high prediction quality. The R-squared score was used to evaluate the model. The LSTM model boasts high prediction accuracy and a low error rate. The locations where crimes are likely to occur within a specific time were identified, and the crime scenes were then identified. Table I summarizes a comparison between related work.

We see from previous studies that the focus was on finding relationships between the time and place of the crime, but the focus was not directly on predicting the number of crimes in a certain month or a certain year. This was the result of this research, which is making a prediction of the number of crimes in 2025 and 2026, which helps the security authorities to take the necessary security measures.

Table 1. Comparison between related work

Authors	Dataset used	Methods	Results
JVimala Devi, and Dr K S Kavitha (2021)	Real data for Sacramento	Use NBeats recurrent neural networks (RNNs) for time-series prediction.	The RNN-LSTM model implemented displayed better performance based on the MAE (Mean Absolute Error)measure.
Wajiha Safat et al. (2021)	Chicago and Los Angeles	Different machine learning algorithms were used: logistic regression, multilayer perceptron (MLP), naive Bayes, k-nearest neighbor (KNN), Extreme gradient boosting (XGBoost),support vector machine (SVM), decision tree, random forest, and time-series analysis	A decrease in crime rates was predicted for Chicago, and a slight increase in crime rates for Los Angeles.
Revanth Sankul et al (2025)	Chicago, New York, Los Angeles Crime	Use Random Forest Classifier	The project successfully created a predictive model that can estimate the likelihood of arrests based on different criminal variables by utilizing historical crime data.
Salah El-Din Ibrahim and Prof. Christina Albert	City of Chicago	set of algorithms, including Random Forest, SVM, K-NN, and MLP.	The SVM algorithm is the best algorithm for achieving the highest accuracy in

Authors	Dataset used	Methods	Results
Reyad (2023) Tsion Eshetu Meskela et al. (2020)			the shortest time(with an accuracy of 0.999976 and a time of 1100 seconds).
Tsion Eshetu Meskela et al (2020)	Bole Sub-City, Addis Ababa, Ethiopia	Using a neural network model for time-series data, the Long Short-Term Memory (LSTM) network.	The LSTM model boasts high prediction accuracy and a low error rate.

3. MATERIALS AND METHOD

This section describes the crime dataset used, the tools used, the research methodology used in this study, and the results we obtained.

3.1 Dataset Description

The crime dataset used in the study is for the city of Los Angeles, and it is acquired from (https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/data) [8]. This website is for the Los Angeles Police Department (LAPD). This dataset reflects incidents of crime in the City of Los Angeles from 2020 to February 2025. This data is transcribed from original crime reports that are typed on paper. It contains 27 attributes (DR_NO, Date Rptd, DATE OCC, TIME OCC, AREA, AREA NAME, Rpt Dist No, Crm Cd, Crm Cd Desc, Mocodes, Vict Age, Vict Sex, Vict Descent, Premis Cd, Part 1-2, Weapon Used Cd, Weapon Desc, Status, Status Desc, Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4, LOCATION, Cross Street, LAT, LON) as shown in Table II. There is another dataset in this study, which is for the unemployment rate and inflation rate in Los Angeles. This dataset is acquired from (https://fred.stlouisfed.org/series/CALOSA7URN) [9],
Unemployed persons are all persons who had no employment during the reference week, were available for work, except for temporary illness, and had made specific efforts to find employment sometime during the 4 week-period ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.
The unemployment rate is the percentage of the civilian labor force who are unemployed [100 times (unemployed/civilian labor force)].(https://www.usinflationcalculator.com/inflation/inflation-in-los-angeles-long-beach-and-anaheim-metropolitan-area/) [10],and(https://www.bls.gov/regions/west/news-release/consumerpriceindex_losangeles.htm)[11]websites. This dataset contains 3three attributes (Data, Inflation Rate, Unemployment Rate) as shown in Table 2.

Table2. Los Angeles Crime Dataset Description.

Name	Type	Description
DR_NO	Text	Division of Records Number: Official file number made up of a 2-digit year, area ID, and 5 digits
Date Rptd	Floating Timestamp	MM/DD/YYYY
DATE OCC	Floating Timestamp	MM/DD/YYYY
TIME OCC	Text	In 24-hour military time.
AREA	Text	The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. These Geographic Areas are sequentially numbered from 1-21.
AREA NAME	Text	The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible for. For example 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles.
Rpt Dist No	Text	A four-digit code that represents a sub-area within a Geographic Area.
Part 1-2	Number	Indicates the crime committed. (Same as Crime Code 1)
Crm Cd Desc	Text	Defines the Crime Code provided.
Mocodes	Text	Modus Operandi: Activities associated with the suspect in the commission of the crime.
Vict Age	Text	determine victim age Two-character numeric
Vict Sex	Text	determine victim sex F - Female M - Male X - Unknown
Vict Descent	Text	Descent Code: A - Other Asian B - Black C - Chinese D - Cambodian F - Filipino G - Guamanian H - Hispanic/Latin/Mexican I - American Indian/Alaskan Native J - Japanese K - Korean L - Laotian O - Other P - Pacific Islander S - Samoan U - Hawaiian V - Vietnamese W - White X - Unknown Z - Asian Indian

Name	Type	Description
Premis Cd	Number	The type of structure, vehicle, or location where the crime took place.
Premis Desc	Text	Defines the Premise Code provided.
Weapon Used Cd	Text	The type of weapon used in the crime
Weapon Desc	Text	Defines the Weapon Used Code provided
Status	Text	Status of the case. (IC is the default)
Status Desc	Text	Defines the Status Code provided.
Crm Cd 1	Text	Indicates the crime committed. Crime Code 1 is the primary and most serious one. Crimes Code 2, 3, and 4 are respectively less serious offenses. Lower crime class numbers are more serious
Crm Cd 2	Text	May contain a code for an additional crime, less serious than Crime Code 1.
Crm Cd 3	Text	May contain a code for an additional crime, less serious than Crime Code 1.
Crm Cd 4	Text	May contain a code for an additional crime, less serious than Crime Code 1.
LOCATION	Text	Street address of crime incident rounded to the nearest hundred block to maintain anonymity.
Cross Street	Text	Cross Street of rounded Address
LAT	Number	Latitude
LON	Number	Longtitude

Table3. Los Angeles inflation and unemployment rate

Column Name	Type	Description
Data	Number	YYYY/DD
Inflation Rate	Number	Represent the inflation rate in Los Angeles
unemployment Rate	Number	Represent the unemployment rate in Los Angeles

3.2 Tool Used

The software and platform that are used in the implementation are listed as follows. Anaconda Python environment for running Python code. Platform used is (Windows 8.1 Pro, processor: Intel® Core™ i5-5200u CPU @ 2.20GHz, installed memory (RAM): 6.00GB, system type:64-bit operating system).

3.3 Research Methodology

The research methodology for analyzing the Los Angeles dataset and studying the impact of specific factors on the crime rate is divided into four distinct phases. Each phase incorporates mathematical and statistical models for data analysis and prediction. Furthermore, each phase is characterized by its unique findings and the purpose for which it is applied to the data. The flowchart of the proposed methodology is shown in Figure 1.

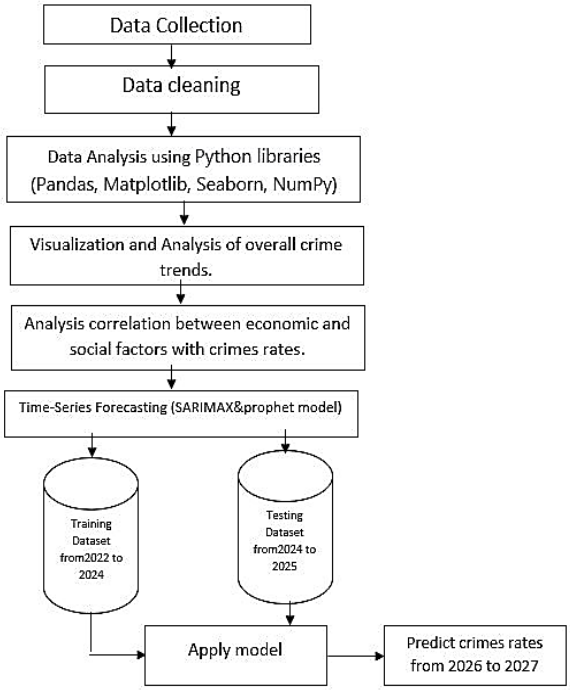

Figure 1: Proposed Flow Chart Of Our Study.

The first stage is to clean and prepare the crime dataset to facilitate analysis. While this stage is not a mathematical model, it is essential for preparing the data for mathematical modeling. The steps of this stage involve data processing and transformation, paving the way for quantitative analysis. These steps are: a. Checking for zero values in the data. b. Processing the victim's gender by adding x to all unknown values to form (female-female, male-male, x-unknown). c. Converting the time to an hour-minute format. d. Creating columns for year, month, and year-month for chronological analysis. The second stage is the analysis of the crime dataset based on several factors such as time, region, victim's age, and gender. This stage focuses on Exploratory Data Analysis (EDA) to understand crime trends based on various factors. a. Trends over time: Analyzing crime counts by year, month, and day involves compiling and plotting time series data. This is a statistical summary. b. Analysis by region: Identifying areas with the highest crime rates. This involves counting and comparing data. c. Victim age and gender: Analyzing crime rates based on the victims' demographics. This includes statistical distribution analysis. The next stage involves studying the impact of economic and social factors and major events on crime rates. This stage examines the relationship between two external factors (the inflation rate and the unemployment rate) and crime rates. To analyze correlations and make time-based comparisons, mathematical/statistical models are applied. Socioeconomic factors are used to explore correlation and make time-based comparisons. A. Social Factors: This involved examining the number of crimes during significant events such as election

periods, the George Floyd protests, the minimum wage for migrant workers, and the COVID-19 pandemic. This included a statistical comparison of crime rates over specific time periods. B. Economic Factors: This involved analyzing the relationship between the inflation rate and the crime rate, and between the unemployment rate and the crime rate. The calculations revealed a weak positive correlation (0.2) between the inflation rate and the crime rate, and no correlation (-0.053) between the unemployment rate and the crime rate. These correlation coefficients are direct outputs of the statistical (mathematical) models. The final stage involves predicting future crime trends using time series forecasting. The SARIMA (Seasonal ARIMA) and prophet models are used for time series forecasting. These are well-established statistical models designed to analyze autocorrelations in data and predict future values based on past observations, particularly when data exhibit seasonality. The models were trained using historical data from 2020 to February 2025. The result of this stage was the prediction of future crime rates for 2026 and 2027. Finally, its implementation relies heavily on statistical analysis, descriptive statistics, correlation analysis, and advanced time-series forecasting models such as SARIMA and Prophet. These models represent the mathematical modeling aspect of the study, enabling quantitative measurement of trends.

4. MATHEMATICAL MODELING FRAMEWORK

Mathematical modeling is the process of using mathematical concepts, equations, and data to create representations of real-world phenomena. These models help us describe, understand, predict, and control various systems, from the physical and biological to the social and economic. They serve as a bridge between theoretical knowledge and practical applications. [19]

This study employs a structured methodological framework, divided into four distinct phases, each incorporating mathematical and statistical modeling techniques for analyzing and predicting crime data. Let's represent the crime dataset using Equation 1 as follows:

$$D = \{ f(y_i, v_i, t_i, r_i, s_i, e_i)\} \text{ where } 0 < i \leq N \quad (1)$$

Where:

D: real data set for Los Angeles.

yi: crime type for record i.

vi: victim attributes (e.g., sex, age).

ti: timestamp of the crime occurrence.

ri: region/location of the crime.

si: social factors or event indicators (e.g., election,protest, Covid).

ei: economic attributes at time t_i(e.g., inflation rate, unemployment rate).

N: total number of records.

Data Preprocessing:

This stage includes the main steps of data preprocessing, i.e., handling missing values, data transformation (scaling, normalization), temporal decomposition, and feature engineering. Let's represent the data transformation crime dataset using Equation 2 as follows:

$$D = T(D) \quad (2)$$

where T includes cleaning missing values, temporal decomposition, and feature engineering.

From (2), we cleaned and prepared the crime dataset for analysis. This step includes several steps. A. Checking for null values in the data. B. Handling the Vict Sex values by adding x to all unknown values to become (F–Female, M-Male, X– Unknown), we used the (to_replace) function from Python to do that. C. Converting time to hour-minute format, we used the (to_datetime) function from Python to do that. D. *Creating columns for year, month, and year_month for chronological analysis, we used (dt.year) and (dt.month) functions from Python to do that.*

Exploratory Statistical Analysis:

Exploratory data analysis (EDA) is a fundamental step in data analysis, enabling the understanding of key data characteristics, the identification of patterns, and the discovery of relationships between different data components. After addressing the issue of missing data and transforming the data, the main characteristics of the data must be determined by examining the distribution, central tendency, and variance of the variables, and identifying outliers or extreme values. Summary statistics, such as the count, Crime distribution, demographic analysis, and Correlation with Social and Economic Factors of the numerical variables, should also be calculated to help construct a suitable model and appropriate analytical methods. The Crime count over a certain period can be estimated using Equation 3 as follows:

$$C_t = \sum_{i=1}^{n} t_i \quad (3)$$

Where Ct represents the Crime count over time , n represents the crime count over time ti.We used the (groupby) function from Python to group data by year and by month.

The Crime count in a certain region can be estimated as given in Equation 4.

$$C_r = \sum_{i=1}^{n} r_i \quad (4)$$

Where Cr represents the crime distribution by region ri , n represents the crime count over time ti. We used the (value_counts ()) function from Python to count the number of crimes in each region. The Crime type count over a certain period and a certain region can be estimated using Equation 5 as follows:

$$C_y = \sum_{i=1}^{n} y_i \quad (5)$$

Where Cy represents the Demographic analysis for each crime type yi, n represents the crime count over time ti . We used the (value_counts) function from Python to count the number of each crime type.

Correlation with Social and Economic Factors:

The population correlation between crime count and Social and Economic Factors can be estimated using the corr() function in Python is primarily used with the pandas library to compute correlation between variables in the dataset. It calculates the correlation coefficient, a value between -1 and +1 that describes the strength and direction of a relationship between two variables.

We used the (corr_matrix) function from Python to find the relationship between Social, economic factors and crime count.

1. Predictive Time-Series Modeling:

In time series analysis used in statistics and econometrics, autoregressive integrated moving average (ARIMA) and seasonal ARIMA (SARIMA) models are generalizations of the autoregressive moving average (ARMA) model to non-stationary series and periodic variation, respectively. All these models are fitted to time series to better understand them and predict future values. [15] The purpose of these generalizations is to fit the data as well as possible. Specifically, ARMA assumes that the series is stationary, that is, its expected value is constant in time. If instead the series has a trend (but a constant variance/autocovariance), the trend is removed by

"differencing"[15], leaving a stationary series. This operation generalizes ARMA and corresponds to the "integrated" part of ARIMA. Analogously, periodic variation is removed by "seasonal differencing".

Non-seasonal ARIMA models are usually denoted ARIMA (p, d, q) where parameters p, d, q are non-negative integers: p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. Seasonal ARIMA models are usually denoted ARIMA (p, d, q) (P, D, Q) m, where the uppercase P, D, Q are the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model, and m is the number of periods in each season. When two of the parameters are 0, the model may be referred to based on the non-zero parameter, dropping "AR", "I", or "MA" from the acronym. For example, ARIMA (1,0,0) is AR (1), ARIMA (0,1,0) is I (1), and ARIMA (0,0,1) is MA (1). [15].

The Prophet library is an open-source library designed for making forecasts for univariate time series datasets. It is easy to use and is designed to automatically find a good set of hyper parameters for the model in an effort to make skillful forecasts for data with trends and seasonal structure by default. [16].

The models were trained using historical data from 2020 to February 2025. The outcome of this stage was the prediction of future crime rates for the years 2026 and 2027.

5. DATA ANALYSIS

Analysis is crucial for transforming raw data into actionable insights, enabling informed decision-making, identifying trends, and enhancing operational efficiency.

Therefore, we analyzed the data to extract a lot of information, including knowing where crimes occurred, knowing the most common types of crimes, knowing the type of victims and their ages, and then it is possible to predict crimes in the future.

1-From analyzing crime by time (yearly), we note that 2024 is the highest crime count (255106 crimes). However, the dataset for 2025 only includes the months of January and February, so their inclusion has no impact on the total number of crimes by year. As shown in Figure 2.

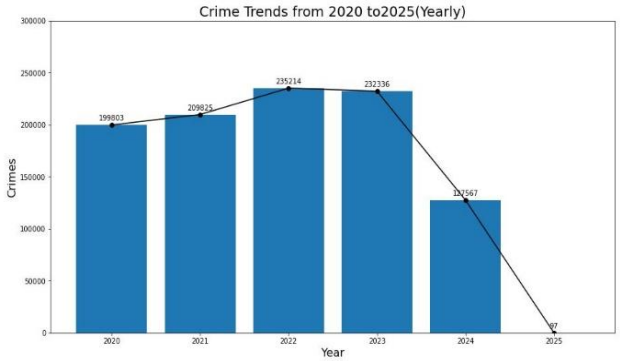

Figure 2: Crime Trends from 2020 to 2025 By Year.

2-From analyzing crime by time (monthly), we note that months ('January', 'February', 'March', 'April') in 2024 are the highest crime count. As shown in Figure 3.

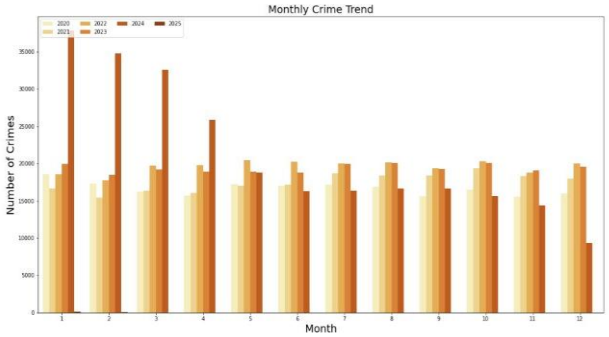

Figure 3: Crime Trends By Month.

3. From analyzing crime by the Average number of crimes per month over the years, we note that months ('January', 'March') have the highest occurrences of crime. As shown in Figure 4.

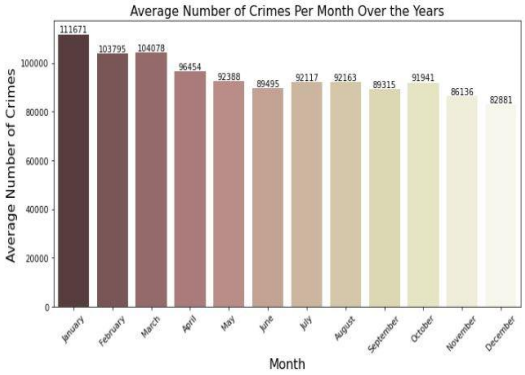

Figure 4: Average Number Of Crimes Per Month Over The Years.

4-From analyzing the type of crimes and their trends over time, we note that stolen vehicles are the most common crime type. As shown in Figure 5.

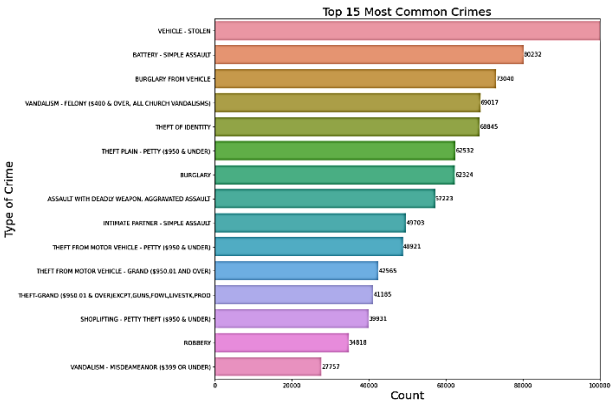

Figure 5: Top 15 Most Common Crimes.

5-From analyzing vehicle-stolen trends over the years from 2020 to 2025, we note that months July and October have the highest occurrences of crime. As shown in Figure 6.

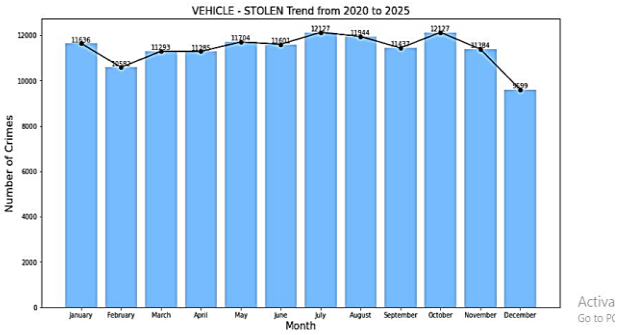

Figure 6: Vehicle-Stolen Trends from 2020 to 2025.

6- From analyzing crimes by regions, we note that the Central area recorded the highest crime incidence. As shown in Figure 7.

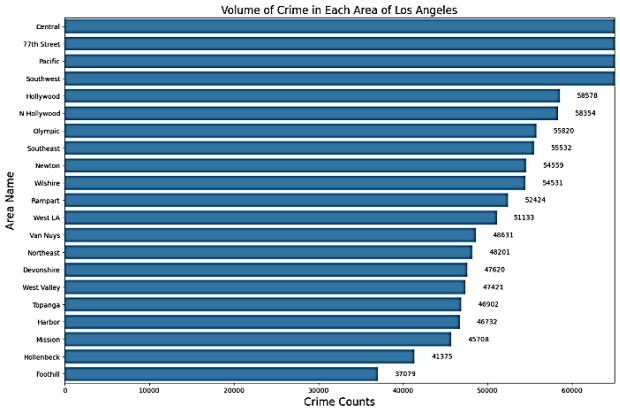

Figure 7: Number Of Crimes In Each Area Of Los Angeles.

7- From analyzing crimes by victim age, we found that the highest rate of committing crimes was among victims aged from 20 to 40. As shown in Figure 8.

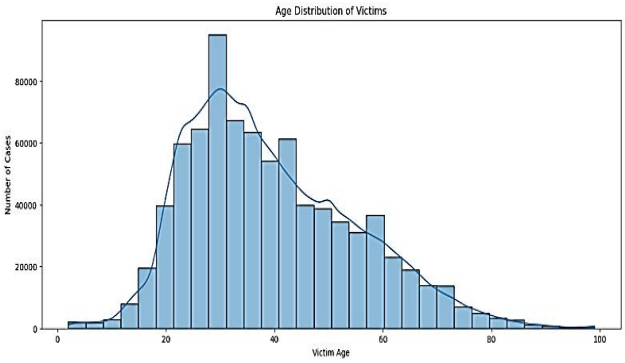

Figure 8: Distribution Of Victims' Age.

8- From analyzing crimes by victim sex, it was found that the highest rate of committing crimes was among male victims.As shown in Figure 9.

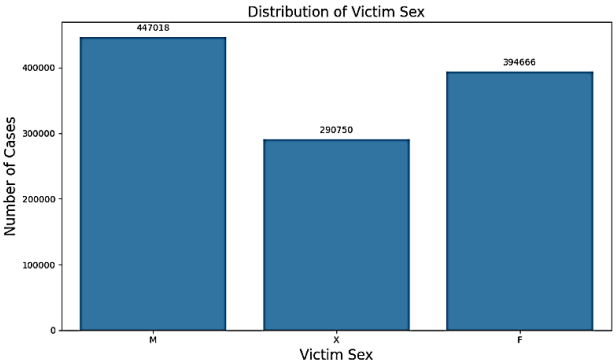

Figure 9: Distribution Of Victims' Sex

In the third phase, we study the impact of economic and social factors and major events on crime rates. Firstly, we study social factors and major events. Many social factors influence crime rates in Los Angeles. Several of these factors were identified during the same period under study. We mention some of these factors, such as 1. US Election Period (from 1/11/2020 to 20/1/2021). 2. George Floyd Protest (from 25/5/2020 to 7/6/2020). 3. Min Wages for selected immigrant workers (from 25/5/2022 to 10/6/2022). 4. COVID (from 1/3/2020 to 25/4/2022). We begin with the first social factor US Election Period. This period was divided into four periods, which are (start of election, formal voting, result day, and the new president inaugurated). We note that the result day has the highest occurrence of the crime rate. As shown in Figure 10.

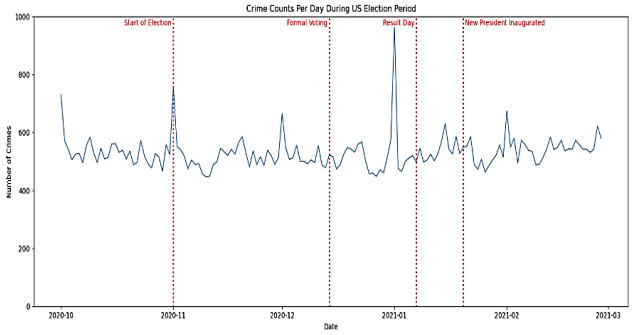

Figure 10: Crime Counts During Election Period.

The second factor is the George Floyd Protest (The George Floyd protests were a series of protests and demonstrations against police brutality that began in Minneapolis in the United States on May 26, 2020). This period was divided into three periods, which are (death-start-end). we note that the highest occurrence of crime was at the start of the George Floyd protests. As shown in Figure 11.

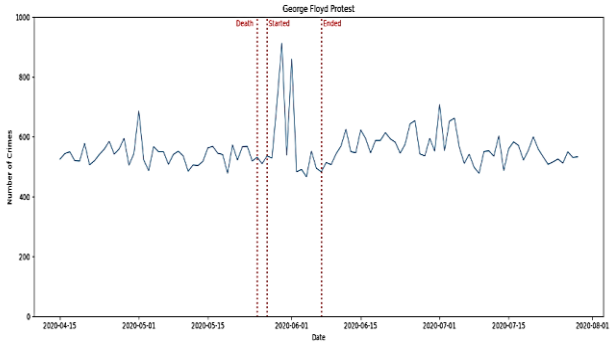

Figure 11: Crime Counts During George Floyd Protests.

The third factor is Min Wages for selected immigrant workers. We note a high occurrence of crime at the start of Min Wages on 28/5/2022. As shown in Figure 12.

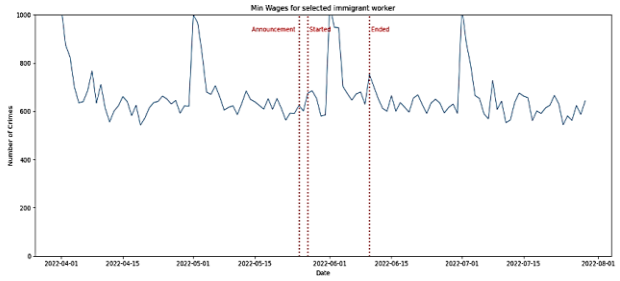

Figure 12: Crime Counts During Min Wages

The fourth factor is COVID [12] (from 1/3/2020 to 25/4/2022). This period was divided into three waves. The first wave (from 1/3/2020 to 1/8/2020).The second wave(from 30/6/2021 to 26/10/2021).The third wave(from 5/1/2022 to 25/4/2022). We note that the highest occurrence of crime is in the third wave. As shown in Figure 13.

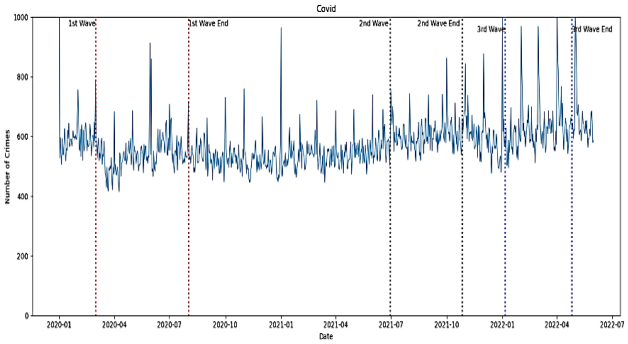

Figure 13: Crime Counts During COVID

In the second part of the third phase, we study the impact of economic factors on crime rates. We study two important economic factors(inflation rate, unemployment rate). We begin with the inflation rate. The date was extracted from the primary data, along with the number of crimes committed. We then linked this data to other data regarding unemployment and inflation rates. The relationship between the date and the inflation rate was analyzed and extracted. As shown in Figure 14.

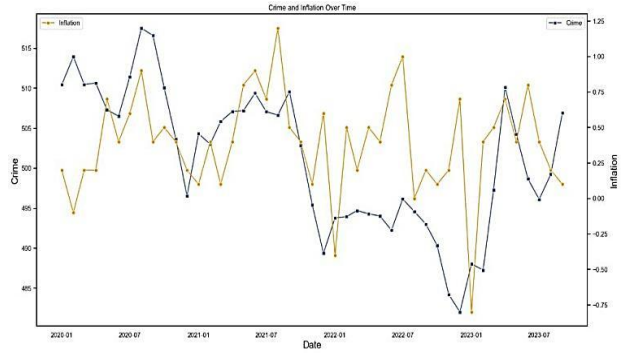

Figure 14: Crime Counts and Inflation Rate

From the Corr_Matrix, we calculate the relation between the inflation rate and the crime rate. We note that there is a weak positive correlation (0.2) between the inflation rate and the crime rate. As shown in Figure 15.

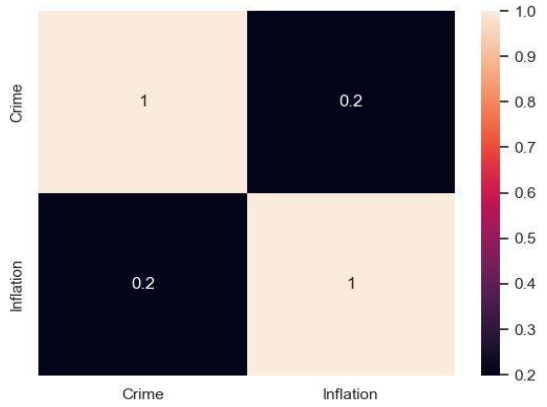

Figure 15: Corr_Matrix for Crime Counts and Inflation Rate

We then examine the impact of the unemployment rate. The relationship between the date and the unemployment rate was analyzed and extracted. As shown in Figure 16.

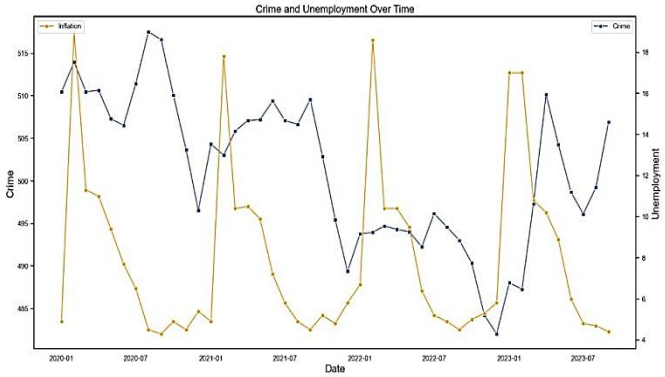

Figure 16: Crime Counts And Unemployment Rate

From the Corr_Matrix, we calculate the relation between the unemployment rate and the crime rate. We note that there is no correlation (-0.053) between the Unemployment rate and the crime rate. As shown in Figure 17.

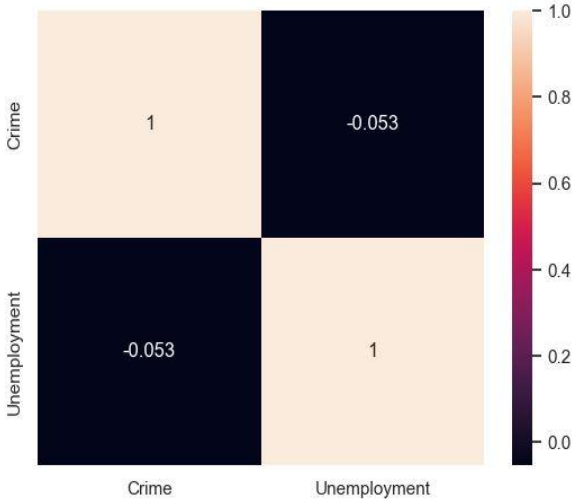

Figure 17: Corr_Matrix for Crime Counts and Unemployment Rate

6. RESULTS AND DISCUSSIONS

In the fourth phase, we used time-series forecasting to predict future crime trends. Time-series forecasting is a statistical method used to analyze and predict future crime trends based on historical

crime data. Analyzing patterns and trends in previous crime incidents aids law enforcement agencies in anticipating and potentially reducing future criminal activity. There are many methods used in time-series forecasting, like ARIMA (AutoRegressive Integrated Moving Average), SARIMA(Seasonal ARIMA), Machine learning, and Prophet.

ARIMA (AutoRegressive Integrated Moving Average). This statistical model analyzes autocorrelations in the data to predict future values based on past observations. We used SARIMA (Seasonal ARIMA), an extension of ARIMA, when the data exhibit seasonality (e.g., more crime during summer months). We trained the model on the past crime data from 2022 to 2024, and we began to test the model from 2024 to 2025and we got this result. As shown in Figure 18.

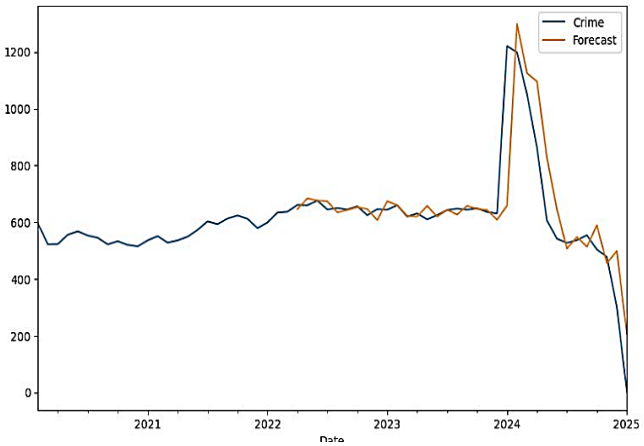

Figure. 18: SARIMA Time-Series Forecasting Learning Model

After training the model, we attempted to predict the future crime counts. We obtained these results for the years 2026 and 2027. We noted that the crime count is expected to decrease in 2026 and 2027 compared to 2024. As shown in Figure 19.

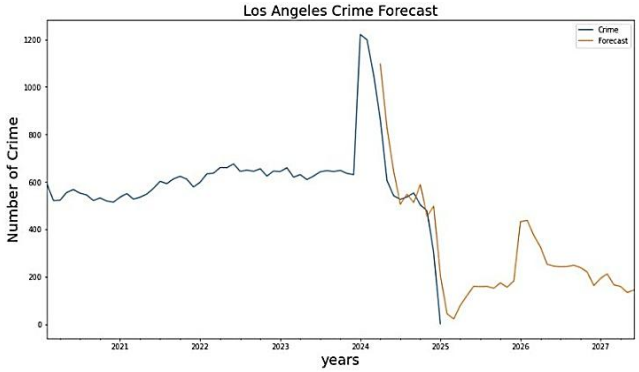

Figure 19: SARIMA Time-Series Forecasting Prediction Model

We also used the Prophet time series forecasting model to predict future crimes. We obtained these results for the years 2026 and 2027. As shown in Figure 20.

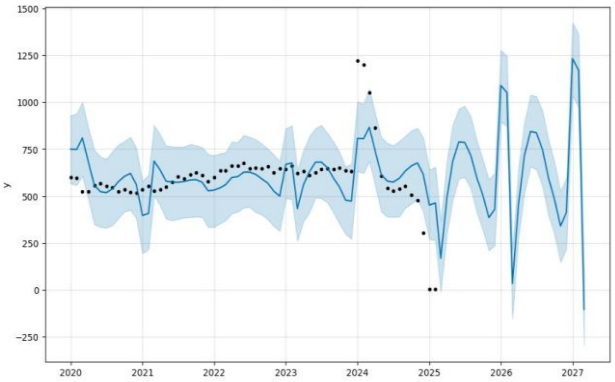

Figure 20: Prophet Time-Series Forecasting Prediction Model

Table 4. Comparison between Prophet and SARIMA.

models	RMSE	MAE	MAPE
SARIMA	3.3166	3.0	15.5555%
Prophet	774.15	411.02	76.33%

We compared two models (Prophet and SARIMA) through three evaluation metrics (RMSE, MAE, and MAPE).
Root Mean Square Error (RMSE) measures the average magnitude of forecasting errors, calculated as the square root of the average squared differences between predicted and observed values. A lower RMSE indicates better model performance, as it signifies smaller discrepancies between the fitted model and actual time series data.

Mean Absolute Error (MAE) measures the average magnitude of errors between predicted and actual values. It serves as a key performance metric, with lower values indicating higher predictive accuracy.

The Mean Absolute Percentage Error (MAPE) calculates the average absolute error as a percentage of the actual values. This makes it a relative measure, independent of the scale of the data.

From the evaluation metrics, we noted that SARIMA is better than Prophet because it has a lower percentage of error than Prophet.

7. CONCLUSION

Crimes represent significant threats to society, national security, and community well-being and must therefore be managed. Crime prediction techniques and the use of historical crime data have become the dominant trend in our society. This study is useful for various agencies and police departments to prevent future crimes. In this study, a historical dataset for the city of Los Angeles was analyzed. The study succeeded in extracting several results that are indicators for predicting future crimes. Among those results was that 2024 had the highest crime count (255106 crimes). The stolen vehicles are the most common crime type. Months July and October have the highest occurrences of crime. The Central region has the highest incidence of crime. There is a weak positive correlation (0.2) between the inflation rate and the crime rate. Finally, we use time-series forecasting to predict future crime trends. After we trained the model using historical data, we attempted to predict the future crime counts. We predicted the crime count for the years 2026 and 2027. According to the time-series forecasting model SARIMA (Seasonal ARIMA), the crime count will decrease compared to 2024. According to the time-series forecasting model prophet crime rates will decrease in the year 2026 compared to 2027. From comparing two time series models, we note that Seasonal ARIMA has better results than Prophet because it has a lower error rate.

8. REFERENCES

[1] Shiju Sathyadevan, Devan M.S., Surya Gangadharan. S.August 2014.Crime Analysis and Prediction Using Data Mining.

[2] Shanjana A.S, 2Dr.R.Porkodi . 2 February 2021.CRIME ANALYSIS AND PREDICTION USING DATA MINING: A REVIEW.

[3] J Vimala Devi, and Dr K S Kavitha.2021.Automating time-series Forecasting on Crime Data using RNN-LSTM.

[4] WAJIHA SAFAT, SOHAIL ASGHAR, and SAIRA ANDLEEB GILLANI .2021. Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques.

[5] Revanth Sankul, Tejaswi Reddy Aruva, Sai Varun Kankal, Greeshma Arrapogula, and ShoeibKhanMohammed. , 2025.Crime data analysis and prediction of arrest using machine learning.

[6] Salah El-Din Abd El-Mohaimen Ibrahim, Prof. Christina Albert Reyad. 2023. A Proposed Big Data Analytics Model for Crimes Predication based on Spatial and Temporal Criminal Hotspot.

[7] Tsion Eshetu Meskela, Yidnekachew Kibru Afework, Nigus Asres Ayele, Muluken Wendwosen Teferi, Tagele Berihun Mengist. 2020. Designing time-series Crime Prediction Model using Long Short-Term Memory Recurrent Neural Network.

[8] Data.world[online](https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/data).

[9] Data.world[online]

(https://fred.stlouisfed.org/series/CALOSA7URN).

[10] Data.world[online] (https://www.usinflationcalculator.com/inflation/inflation-in-los-angeles-long-beach-and-anaheim-metropolitan-area/).

[11] Data.world[online] (https://www.bls.gov/regions/west/news-release/consumerpriceindex_losangeles.htm).

[12] Shelby M. Scott, Louis J. Gross. 2021. COVID-19 and crime: Analysis of crime dynamics amidst social distancing protocols.

[13] H. Benjamin Fredrick David1 and A. Suruliandi. 2017 SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES.

[14] Alkesh Bharati1, Dr Sarvanaguru RA.K2. 2018 Crime Prediction and Analysis Using Machine Learning.

[15] Rob J Hyndman and George Athanasopoulos. 2018.Forecasting: Principles and Practice.

[16] Jason Brownlee. 2020.Time Series Forecasting With Prophet in Python.