

Patient Disabilities: Machine learning for Crime rates

DAB322A Capstone Project Group 7

Project Abstract

Abhishek (0803640)

Keyur Bhalala :(0805748)

Komal sharma :(0891433)

St Clair, Mississauga

Prof. Savita Seharawat

July 26, 2023

Abstract

This study examines the trends in crime rate from the year 2001 to the present day, aiming to shed light on the evolving patterns of criminal activities over the past two decades. The primary objective is to analyze and interpret the data to gain insights into the fluctuations and long-term changes in criminal incidents during this period.

The data utilized in this research comprise official crime statistics collected from various law enforcement agencies and relevant government sources. The study adopts a quantitative approach, employing statistical methods and data visualization techniques to explore the crime rate trends across different categories of offenses, such as violent crimes, property crimes, and white-collar crimes.

The findings of this study reveal notable shifts in crime rates over the years, highlighting significant variations in crime patterns across regions and crime types. By examining the factors that might have contributed to these changes, this study aims to provide a deeper understanding of the complexities surrounding crime dynamics.

Keywords: Machine learning, classification, predictive modeling, cross validation, k-Means clustering.

Objectives/research questions:

- 1. How many crimes we can happen in future?**
- 2. Which region will get most of the crimes in the future?**
- 3. Which area is having most of the crimes?**

4. What is the status of the emergency crime, like whether it is solved or not?

5. What are the most common types of incidents reported in the dataset? Is there any same type of pattern?

Tools:

We will explore the dataset using excel, furthermore, we used the python for data preprocessing EDA and data modeling

Introduction:

In this project we are looking to analyze the different disabilities estimates. We would find the growing field that utilizes advanced data analytics and machine learning techniques to extract valuable insights from the crime's data and enhance public security.

Theme of my dataset:

According to the dataset which we have we can perform Classification because most of our data types are categorical. Based on the categorical values so we can apply classification.

- Classification
- Python: For Machine learning
- Tableau: For Data visualization
- Word: For Abstract and reports

Literature review:

Methodology: This research focuses on utilizing Exploratory Data Analysis (EDA) and Data Modeling techniques to process and analyze crime rate data from the year 2001 to the present. The primary objective is to gain insights into the temporal patterns and underlying factors influencing crime rates over the years, ultimately enabling better-informed decision-making for crime prevention and public safety strategies.

Data Description: The dataset used in this study encompasses crime rate statistics for a specific geographical area, covering the period from 2001 to the present. The data is sourced from reliable law enforcement and government databases and is structured to include information on different types of crimes, geographical locations, and timestamps.

Exploratory Data Analysis (EDA): The EDA phase involves comprehensive data exploration to uncover patterns, trends, and potential anomalies. Descriptive statistics, data visualization, and other EDA techniques will be employed to identify the distribution of crime rates across time, regions, and crime categories. The EDA process will also examine relationships between crime rates and various socio-economic variables, such as population density, unemployment rates, and education levels.

Data Modeling:

For data modeling, advanced statistical and machine learning algorithms will be applied to develop predictive models. Time series analysis will be employed to forecast future crime rates

based on historical patterns, enabling the identification of potential hotspots and proactive allocation of resources. Additionally, machine learning models, such as regression, decision trees, and clustering, will be utilized to uncover hidden patterns and correlations within the data, offering valuable insights for crime prevention strategies.

Expected Outcomes: Identification of Temporal Patterns: The EDA process is expected to reveal trends and seasonality in crime rates over the years, aiding in the understanding of cyclical patterns and long-term shifts.

Geographical Hotspot Analysis: Data modeling will assist in identifying crime hotspots, allowing law enforcement agencies to focus resources on areas with higher crime rates.

Socio-Economic Factors Impact: By analyzing the relationships between crime rates and socio-economic variables, this study aims to determine the factors influencing crime trends in the region.

Predictive Insights: Time series forecasting and machine learning models are expected to provide predictive insights, empowering proactive crime prevention measures.

Data Dictionary:

Column Name	Description	Type
Rowid	Unique row identifier	categorical

LocationAbbr	it represents the location name in short form	categorical
LocationDesc	Location abbreviation	categorical
DataSource	Location description	categorical
Category	Abbreviation of data source name	categorical
Indicator	Category description	categorical
Response	Indicator description	categorical
Data_Value_Unit	Indicator response	categorical
Data_Value_Type	The unit for a data value such as %	categorical
Data_Value_Type	The data type such as age-adjusted prevalence and prevalence	categorical
StratificationCategory1	<div> <div>Case-Number</div> <div>object</div> </div> <div> <div>Date</div> <div>object</div> </div> <div> <div>Block</div> <div>object</div> </div> <div> <div>IUCR</div> <div>object</div> </div> <div> <div>Primary-Type</div> <div>object</div> </div> <div> <div>Description</div> <div></div> </div> <div> <div>Case-Number</div> <div>object</div> </div> <div> <div>Date</div> <div>object</div> </div> <div> <div>Block</div> <div>object</div> </div> <div> <div>IUCR</div> <div>object</div> </div> <div> <div>Primary-Type</div> <div>object</div> </div> <div> <div>Description</div> <div></div> </div>	categorical
Stratification1	<div> <div>Description</div> <div>object</div> </div> <div> <div>Location-Description</div> <div>object</div> </div> <div> <div>Arrest</div> <div>bool</div> </div> <div> <div>Domestic</div> <div>bool</div> </div> <div> <div>Beat</div> <div>int64</div> </div> <div> <div>District</div> <div>float64</div> </div> <div> <div>Ward</div> <div>float64</div> </div> <div> <div>Community-Area</div> <div>float64</div> </div> <div> <div>FBI Code</div> <div>object</div> </div> <div> <div>X-Coordinate</div> <div>float64</div> </div> <div> <div>Y-Coordinate</div> <div>float64</div> </div> <div> <div>Year</div> <div>int64</div> </div> <div> <div>Updated-On</div> <div>object</div> </div> <div> <div>Latitude</div> <div>float64</div> </div> <div> <div>Longitude</div> <div>float64</div> </div> <div> <div>Location</div> <div>object</div> </div> <div> <div>dtype: object</div> <div></div> </div>	categorical

StratificationCategory2	Community-Area float64 FBI Code object X-Coordinate float64 Y-Coordinate float64 Year int64 Updated-On object Latitude float64 Longitude float64 Location object dtype: object	categorical
Stratification2	X-Coordinate float64 Y-Coordinate float64 Year int64 Updated-On object Latitude float64 Longitude float64 Location object dtype: object	categorical
CategoryID	Category identifier	categorical
IndicatorID	Indicator Identifier	categorical
LocationID	Location Identifier	categorical
ResponseID	Response Identifier	categorical
DataValueTypeID	Identifier for the Data Value Type	categorical
StratificationCategoryID1	Stratification Category 1 Identifier	categorical
StratificationID1	Stratification 1 Identifier	categorical
StratificationCategoryID2	Stratification Category 2 Identifier	categorical
StratificationID2	Stratification 2 Identifier	categorical

Data preprocessing:

Data preprocessing is a crucial step in the data analysis and machine learning workflow. It involves transforming raw data into a format suitable for analysis or modeling, ensuring that the

data is accurate, consistent, and ready for further processing. Proper data preprocessing can significantly impact the performance and accuracy of machine learning models. Here are some common techniques used in data preprocessing:

Missing Values:

Attribute	No. of missing values
Data_Value	78283
Low_Confidence_Limit	78283
High_Confidence_Limit	78283
Number	78283

- **Analyzing the dataset reveals that Location Description has most of the null values for the criminal-type of 'Deception'. So, the rows containing 'Description' as - Deception are specified and will be filled with 'others' and 'most frequent entries' in two different parts and will accept the most apt set of observations.**

ID	Case-Number	Date	Block	IUCR	Primary-Type	Description	Location-Description	Arrest	Domestic	...	Ward	Community-Area	FBI Code	X-Coordinate	Y-Coordinate	Year	Updated-On	Latitude	Longitude	Location	
0	10224738	HY411648	09/05/2015 01:30:00 PM	043XXS WOODST	486	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENCE	False	True	...	12.0	61.0	08B	1165074.0	1875917.0	2015	02/10/2018 03:50:01 PM	41.815117	-87.6700	(41.815117282, -87.66999562)
1	10224739	HY411615	09/04/2015 11:30:00 AM	008XXN CENTRAL AVE	870	THEFT	POCKET-PICKING	CTA BUS	False	False	...	29.0	25.0	6	1138875.0	1904869.0	2015	02/10/2018 03:50:01 PM	41.895080	-87.7654	(41.895080471, -87.765400451)

2 rows x 22 columns

● Select rows where 'Column' has specific entries.

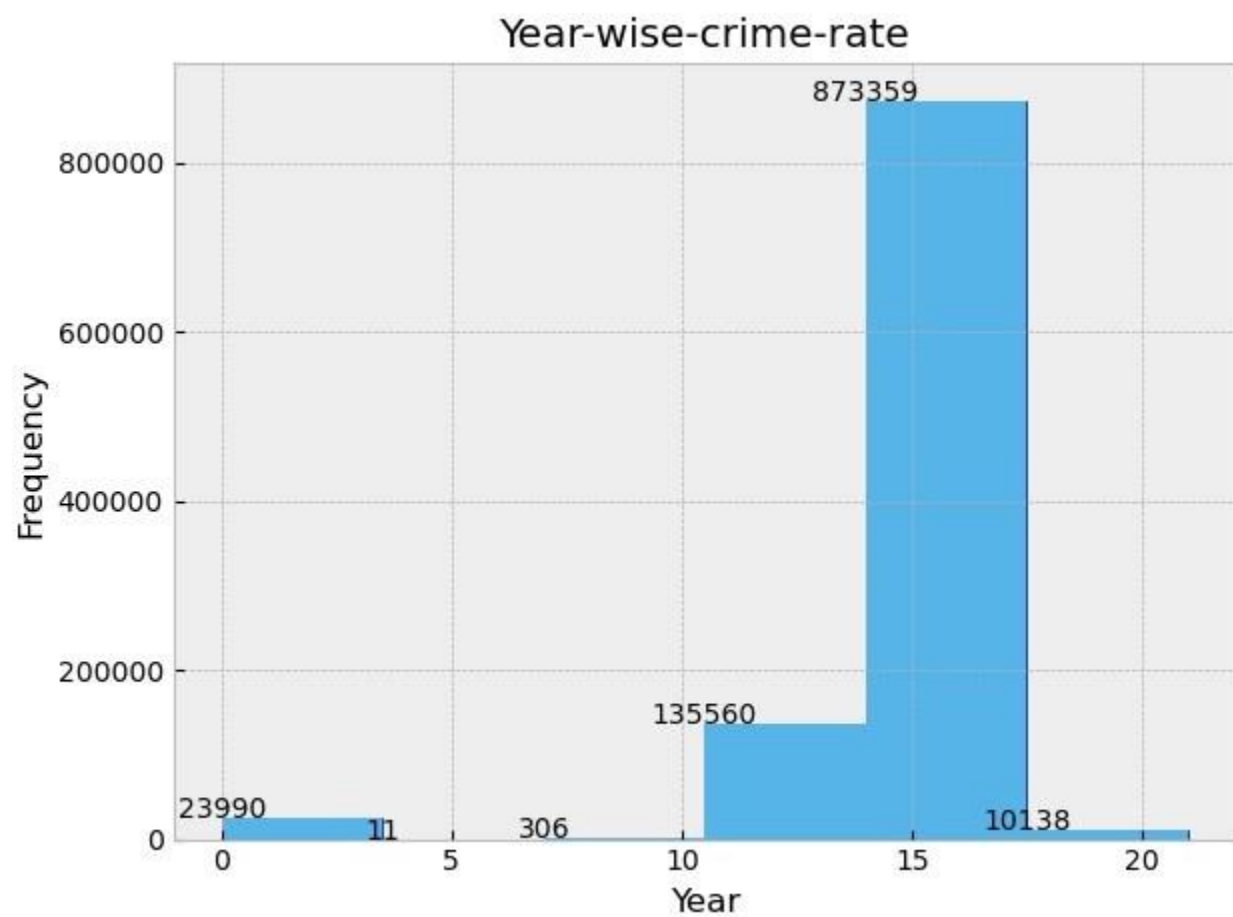
ID	Case-Number	Date	Block	IUCR	Primary-Type	Description	Location-Description	Arrest	Domestic	...	World	Community-Area	FBI Code	X-Coordinate	Y-Coordinate	Year	Updated-On	Latitude	Longitude	Location	
9	11645836	JC212333	05/01/2016 12:25:00 AM	055X X S ROCKWELL ST	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER \$ 300	OTHER	False	False	...	15.0	63.0	11	NaN	NaN	2016	04/06/2019 04:04:43 PM	NaN	NaN	NaN
46	10224790	HY411713	09/01/2015 09:00:00 AM	0000 X W ILLINOIS ST	1122	DECEPTIVE PRACTICE	COUNTERFEIT CHECK	RESTAURANT	False	False	...	42.0	8.0	10	1176121.0	1903605.0	2015	02/10/2018 03:50:01 PM	41.890854	-87.628644	(41.89085, -87.62864, 182)
54	11645527	JC212744	02/02/2015 10:00:00 AM	069X X W ARCHER AVE	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER \$ 300	OTHER	False	False	...	23.0	56.0	11	NaN	NaN	2015	04/06/2019 04:04:43 PM	NaN	NaN	NaN
58	10224806	HY411696	09/03/2015 01:30:00 PM	010X X N HAMLIN AVE	1130	DECEPTIVE PRACTICE	FRAUD OR CONFIDENCE GAME	RESIDENCE	False	False	...	27.0	23.0	11	1150857.0	1906759.0	2015	02/10/2018 03:50:01 PM	41.900041	-87.721344	(41.90004, -87.72134, 753)
69	10224820	HY411701	09/05/2015 12:00:00 PM	010X X N HAMLIN AVE	1130	DECEPTIVE PRACTICE	FRAUD OR CONFIDENCE GAME	APARTMENT	False	False	...	27.0	23.0	11	1150857.0	1906759.0	2015	02/10/2018 03:50:01 PM	41.900041	-87.721344	(41.90004, -87.72134, 753)
...
1048405	1313348	G003830	01/02/2001 07:00:00 PM	001X X N STATE ST	1210	DECEPTIVE PRACTICE	THEFT OF LABOR/SERVICES	CTA PLATFORM	True	False	03:40 PM	...	035	8035491)
1048445	11693789	JC271126	05/20/2019 03:48:00 PM	075X X S DORCHESTER AVE	1150	DECEPTIVE PRACTICE	CREDIT CARD FRAUD	SIDEWALK	True	False	...	8.0	43.0	11	1186848.0	1855324.0	2019	06/26/2019 04:10:52 PM	41.758119	-87.590782	(41.75811, -87.59078, 437)
1048462	1313498	G003618	01/02/2001 04:45:00 PM	021X X N AUSTIN AV	1210	DECEPTIVE PRACTICE	THEFT OF LABOR/SERVICES	RESTAURANT	True	False	...	NaN	NaN	11	1136081.0	1913966.0	2001	08/17/2015 03:03:40 PM	41.920094	-87.775445	(41.92009, -87.77544, 95)

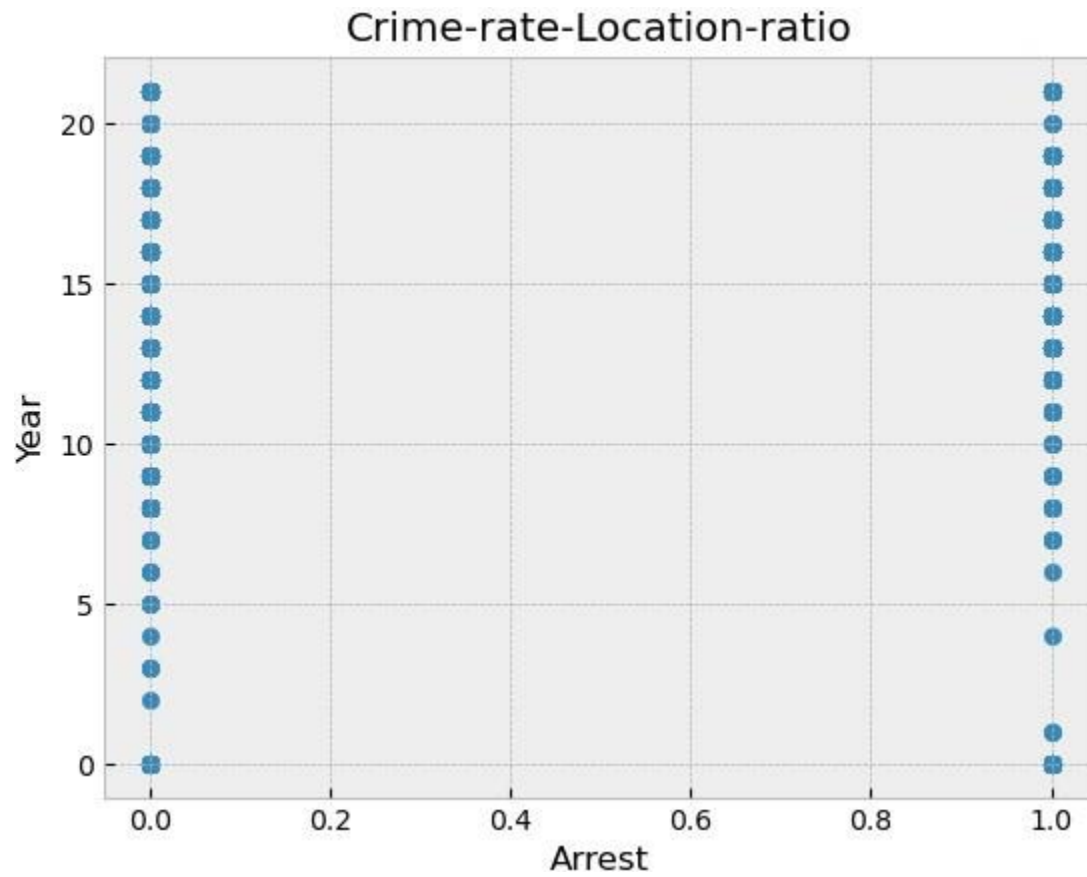
ID	Case-Number	Date	Block	IUCR	Primary-Type	Description	Location-Description	Arrest	Domestic	...	Ward	Community-Area	FBI Code	X-Coordinate	Y-Coordinate	Year	Updated-On	Latitude	Longitude	Location	
1048468	1313512	G004864	01/03/2001 10:00:00 AM	108X X S CAL UME TAV	1130	DECEPTIVE PRACTICE	FRAUD OR CONFIDENCE GAME	RESIDENCE	False	False	...	NaN	NaN	11	1180551.0	1832838.0	2001	08/17/2015 03:03:40 PM	41.696562	-87.614548	(41.696562, -87.614548)
10511	1313632	G001784	01/01/2001 08:05:00 PM	011X X W BRY N MAW RAV	1220	DECEPTIVE PRACTICE	THEFT OF LOST/MISLAI D PROP	ALLEY	True	True	...	NaN	NaN	11	1167657.0	1937313.0	2001	08/17/2015 03:03:40 PM	41.983537	-87.614548	(41.983537, -87.614548)

71397 rows x 22 columns

● In order to detect the outliers in the dataset it is important to sketch the histogram and for the

#package of matplotlib is to be imported

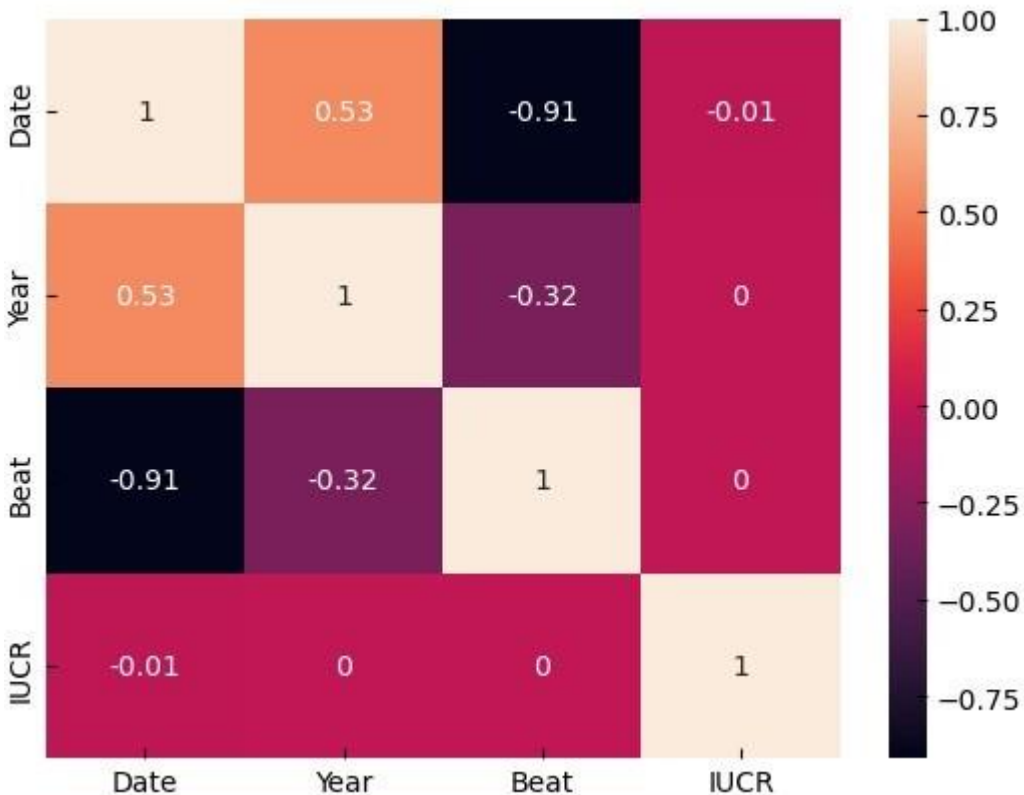




Dividing the data into input and output variables

Date	Year	Beat	
0	839	13	924
1	838	13	1511

- **Analyzing the heatmap for input and output variables**



- **Checking the accuracy of the decision tree model.**

98.47 %

Literature Review:

Crime rate trends have been a subject of great interest for researchers, policymakers, and the general public. Understanding how crime rates have evolved over time is crucial for designing effective crime prevention and law enforcement strategies. This literature review aims to provide a comprehensive overview of the research conducted on crime rates from 2001 to the present, identifying key patterns, factors influencing crime trends, and the effectiveness of crime control measures.

Crime Rate Trends from 2001 to the Present:

Numerous studies have explored the overall trends in crime rates over the past two decades. Initial studies indicate a decline in overall crime rates during the early 2000s, followed by a more nuanced

and complex picture in the latter part of the period. While some regions and countries have continued to experience declining crime rates, others have witnessed fluctuations or even increases in specific crime categories.

Factors Influencing Crime Rate Trends:

Researchers have attempted to identify various factors that influence crime rate trends during this period. Socioeconomic factors, such as poverty, unemployment, and income inequality, have been consistently linked to crime rates. Additionally, demographic shifts, urbanization, and changes in the age structure of populations have also been explored as potential contributors to crime trends.

Technological Advancements and Crime:

The advent of the digital age has brought new challenges and opportunities related to crime. Studies have examined how technology, particularly the internet and social media, has been linked to the emergence of cybercrime and its impact on traditional crime rates. Moreover, advances in surveillance technology and data analysis have influenced crime prevention efforts.

Legislative and Policy Changes:

Changes in legislation and crime control policies have been integral to understanding crime rate trends. Scholars have evaluated the impact of various criminal justice reforms, such as changes in sentencing laws, policing strategies, and community-based initiatives, on crime rates.

Crime and Public Health:

An emerging area of research focuses on the intersection between crime and public health. Studies have explored the connection between crime rates and public health issues, such as substance abuse, mental health, and community well-being, highlighting the importance of a holistic approach to crime prevention.

Conclusion:

Overall, the integration of EDA and Data Modeling in the analysis of crime rate data has proven to be a powerful approach in understanding the complexities of criminal activities. The findings of this research provide valuable knowledge for developing evidence-based strategies, allocating resources efficiently, and fostering community-oriented approaches to crime prevention.

While this study contributes to our understanding of crime trends in the specified region, it is essential to acknowledge the limitations of the data and the models used. As with any data analysis, the accuracy and effectiveness of the conclusions are contingent on data quality and the selected modeling techniques.

