UITS

Future will be better than thy past

An initiative of PHP Family

University of Information Technology & Sciences

A SKELETON WITH INPUT OF ARTICALS FINAL REPORT

Scientific Research and Methodology
Course Code: CSE 418 (Dual)
Course Teacher:

Ratri Datta Lecturer, Department of CSE	Md. Moradul Siddique Lecturer, Department of CSE

Date of Submission: December 18, 2024

Title of the Research: Enhancing Crime Prediction Accuracy Through Machine Learning: A Predictive Policing Approach
Team Name: Digital_Warrior

Team Member:
Md. Shakibul Islam Ramim (2125051063)
Md.Nahian Islam Emon (2125051114)
Fazlay Rabbi (2125051070)
Mst. Sumi Akter (2125051037)

Enhancing Crime Prediction Accuracy Through Machine Learning: A Predictive Policing Approach

Abstract

Crime poses a significant threat to the security and jurisdiction of any nation. Consequently, crime analysis has gained increasing importance as it involves discerning the when and where of criminal activities through the analysis of spatial and temporal data. Traditional methods such as paperwork, reliance on investigative judges, and statistical analysis have proven inefficient in accurately predicting the time and location of crimes. However, the integration of machine learning and data mining techniques into crime analysis has led to a substantial improvement in the accuracy of crime analysis and prediction. This study delves into various aspects of criminal analysis and prediction using a range of machine learning and data mining methods. It aims to provide a succinct overview of how these algorithms are employed in crime prediction, based on the accuracy metrics of previous research. The intention is not only to inform crime researchers about these techniques but also to support future endeavors in refining crime analysis. This review study encompasses an exploration of crime definitions, challenges in prediction systems, and classifications, accompanied by a comparative analysis. Through a comprehensive examination of the literature, it becomes evident that supervised learning approaches have been the predominant choice for crime prediction in numerous studies, surpassing other methodologies. Furthermore, Logistic Regression emerges as the most robust method for predicting crime based on existing research findings.

Keywords: Crime prediction, GIS, Cluster, Data Mining.

INTRODUCTION

Law violations pose a significant threat to the functioning of the justice system and demand effective measures for prevention. Computational crime prediction and forecasting can play a pivotal role in enhancing the safety of urban areas. The complexity of handling vast volumes of intricate data within big data sets makes it challenging to make timely and accurate predictions regarding criminal activities. This presents both challenges and opportunities in the realm of computational crime prediction.

The accuracy of predicting crime rates, types, and high-risk locations based on historical patterns remains a pressing issue. Despite substantial research efforts, there is a persistent need for robust prediction algorithms that can guide law enforcement efforts, particularly in targeting police patrols toward potential criminal events [1].

Crime analysis, as a methodology, is employed to identify areas with high crime incidences, but it is by no means a straightforward process. In 2020, Geographical Information Systems (GIS) emerged as a non-machine learning tool used for analyzing temporal and spatial data. GIS, employing crime hotspot techniques primarily dependent on crime types, aimed to reduce crime rates [2].

Crime rate prediction can be defined as a method to create systems that discern future crime patterns, aiding law enforcement in solving crimes and subsequently reducing crime rates in the real world. On the other hand, crime forecasting involves predicting crimes far into the future, sometimes years ahead, to enhance crime prevention efforts. This can be achieved by utilizing time series approaches to identify future crime trends from time series data.

In the realm of crime analysis within data mining, various methods are employed, including statistical approaches [3] [4] [5], visualization techniques [6] [7] [8], unsupervised learning, and supervised learning methods [9] [10] [11]. Visualization methods encompass presenting connections between geographic views and other crime-related data, such as geographic profiling, GIS-based crime mapping [12] [13] [14], crime prediction, and asymmetric mapping [15] [16] [17]. Additionally, clustering methods, which have gained popularity, are employed to uncover patterns and groups within crime data, contributing to criminal behavior analysis, crime pattern recognition, criminal association analysis, and incident pattern recognition [18] [19] [20].

The development of machine learning algorithms has significantly advanced crime data analysis. These algorithms have been utilized to preprocess and cluster data, extract crime locations from raw data [21], and apply both supervised and unsupervised machine learning models to analyze data patterns based on time and location of crimes, leading to more precise predictions [22]. Furthermore, machine learning algorithms have been instrumental in investigating the factors contributing to crime in specific areas by analyzing historical data collected from previous years in those regions [23].

In recent times, the development of classification algorithms, particularly machine learning algorithms, has further bolstered crime prediction [24]. Researchers have endeavored to correlate time with crime by considering various factors, aiding in the resolution and prevention of crimes. In 2018, Fourier series was proposed as an analytical technique to establish flexible mathematical models for time-periodic effects, demonstrating the effectiveness of analytical techniques in linking time with crime prediction, although its applicability may vary depending on the type of crime [25].

While machine learning algorithms are widely employed in the field of crime prediction, they are not without limitations and do not surpass the utility of data mining techniques, each offering its own performance characteristics and outcomes.

This study's primary objective is to acquaint readers with previous research and the corresponding levels of accuracy achieved, presented in tabular format. Its main contribution lies in presenting applications of machine learning and data mining in crime prediction, categorizing studies based on different techniques, and providing concise overviews of each methodology used for mining crime data. Additionally, the study identifies some challenges faced by developers of such systems. However, there are limitations to the existing body of work, including the lack of extensive geographical coverage, limited generality when applying the same system to different crime datasets, scarcity of studies focusing on predicting criminal actions, and challenges researchers encounter due to missing or duplicated information within online crime datasets.

CRIME DEFINITION AND DESCRIPTION

Law violations pose a significant threat to the functioning of the justice system and demand effective measures for prevention. Computational crime prediction and forecasting can play a pivotal role in enhancing the safety of urban areas. The complexity of handling vast volumes of intricate data within big data sets makes it challenging to make timely and accurate predictions regarding criminal activities. This presents both challenges and opportunities in the realm of computational crime prediction.

The accuracy of predicting crime rates, types, and high-risk locations based on historical patterns remains a pressing issue. Despite substantial research efforts, there is a persistent need for robust prediction algorithms that can guide law enforcement efforts, particularly in targeting police patrols toward potential criminal events [1].

Crime analysis, as a methodology, is employed to identify areas with high crime incidences, but it is by no means a straightforward process. In 2020, Geographical Information Systems (GIS) emerged as a non-machine learning tool used for analyzing temporal and spatial data. GIS, employing crime hotspot techniques primarily dependent on crime types, aimed to reduce crime rates [2].

Crime rate prediction can be defined as a method to create systems that discern future crime patterns, aiding law enforcement in solving crimes and subsequently reducing crime rates in the real world. On the other hand, crime forecasting involves predicting crimes far into the future, sometimes years ahead, to enhance crime prevention efforts. This can be achieved by utilizing time series approaches to identify future crime trends from time series data.

In the realm of crime analysis within data mining, various methods are employed, including statistical approaches [3] [4] [5], visualization techniques [6–8], unsupervised learning, and supervised learning methods [9] [10] [11]. Visualization methods encompass presenting connections between geographic views and other crime-related data, such as geographic profiling, GIS-based crime mapping [12] [13] 14], crime prediction, and asymmetric mapping [15] [16] [17]. Additionally, clustering methods, which have gained popularity, are employed to uncover patterns and groups within crime data, contributing to criminal behavior analysis, crime pattern recognition, criminal association analysis, and incident pattern recognition [18] [19] [20].

The development of machine learning algorithms has significantly advanced crime data analysis. These algorithms have been utilized to preprocess and cluster data, extract crime locations from raw data [21], and apply both supervised and unsupervised machine learning models to analyze data patterns based on time and location of crimes, leading to more precise predictions [22]. Furthermore, machine learning algorithms have been instrumental in investigating the factors contributing to crime in specific areas by analyzing historical data collected from previous years in those regions [23].

In recent times, the development of classification algorithms, particularly machine learning algorithms, has further bolstered crime prediction [24]. Researchers have endeavored to correlate time with crime by considering various factors, aiding in the resolution and prevention of crimes. In 2018, Fourier series was proposed as an analytical technique to establish flexible mathematical models for time-periodic effects, demonstrating the effectiveness of analytical techniques in linking time with crime prediction, although its applicability may vary depending on the type of crime [25].

While machine learning algorithms are widely employed in the field of crime prediction, they are not without limitations and do not surpass the utility of data mining techniques, each offering its own performance characteristics and outcomes.

This study's primary objective is to acquaint readers with previous research and the corresponding levels of accuracy achieved, presented in tabular format. Its main contribution lies in presenting applications of machine learning and data mining in crime prediction, categorizing studies based on different techniques, and providing concise overviews of each methodology used for mining crime data. Additionally, the study identifies some challenges faced by developers of such systems. However, there are limitations to the existing body of work, including the lack of extensive geographical coverage, limited generality when applying the same system to different crime datasets, scarcity of studies focusing on predicting criminal actions, and challenges researchers encounter due to missing or duplicated information within online crime datasets.

CHALLENGES OF PREDICTION SYSTEMS

Researchers and government security agencies encounter several challenges when attempting to predict the location and timing of crimes, as well as in selecting the most effective methods for doing so. Furthermore, computer science researchers employing machine learning, data mining, and spatial-temporal data face their own set of obstacles. In 2012 and 2016, near-repeat-victimization and repeat - victimization methods were introduced to forecast crimes in residential areas, streets, and regions. These methods propose that when a crime occurs in a specific area, there is a significant likelihood of an increased occurrence of other crimes in the same vicinity [27] [28].

Challenges faced by developers of crime prediction systems include:

a. The substantial volume of data necessitates extensive storage capacity.

b. Crime-related data often exist in diverse formats, such as text, images, graphs, audio, relational data, unstructured data, and semi-structured data [29]. Consequently, the process of converting these data into a comprehensible format presents a challenge.

c. In the realm of machine learning, accurately assigning the appropriate label (e.g., prediction

or output) to an instance (e.g., context or input) poses a significant challenge.

d. Selecting the most suitable data mining algorithm that can yield superior results compared to the currently utilized algorithms is another challenge.

e. Environmental and contextual factors, such as the presence or absence of law enforcement and weather conditions, exert an influence on the likelihood of criminal activity.

These factors can lead crime prediction algorithms to make substantial errors. To attain high prediction accuracy, any crime forecasting system must account for these environmental and contextual variations.

RELATED WORK

The advent of extensive data resources has revolutionized the application of machine learning and data mining techniques, providing law enforcement with powerful tools for crime detection and reduction. Proper parameter selection in these techniques enables law enforcement agencies to effectively analyze data, uncover links between criminal activities, and identify patterns and trends, ultimately enhancing their ability to combat criminal activities more efficiently [5].

This section delves into a discussion and analysis of prior research in this domain, which encompasses a wide range of approaches. Some studies focus on crime analysis and prediction, while others apply Artificial Intelligence (AI), machine learning, or data mining, which are subfields of AI, to forecast violent crimes using spatial and temporal data.

During our survey, we identified five significant surveys or overviews related to crime prediction and machine learning or data mining. The earliest one dates to 2011, where various methods were explored for extracting spatial patterns, known as spatial data mining (SDM) algorithms. These methods included co-location mining, spatial clustering, spatial hot spots, spatial outliers, spatial auto-regression, conditional auto-regression, and geographically weighted regression. This survey highlighted the effectiveness of these SDM algorithms and their practical applicability, emphasizing the need for additional methods to validate the hypotheses generated by these algorithms [32].

In 2015, researchers investigated crime prediction using data mining and machine learning techniques. They considered a variety of crime-related variables and found that factors such as age, alcohol consumption, hot spots, media exposure, and certain policies did not significantly affect crime rate predictions.

While the discussion was insightful, the study lacked a comprehensive conclusion [33].

In 2016, another survey analyzed over 100 applications of data mining in the context of crime. Researchers provided a concise summary by presenting a table that listed the techniques used alongside specific software, the relevant study areas, and the expected uses and functions. They recommended improving the utility of data mining techniques in crime data analysis through enhanced training and education [34].

n 2019, a systematic review of crime prediction and data mining studies conducted between 2004 and 2018 classified research works based on the data mining techniques employed. This analysis revealed a common challenge: as datasets grew, the overall performance of the systems decreased. This observation was consistent across 40 covered papers [35]. Finally, in 2020, another systematic review focused on spatial crime forecasting. This study analyzed 32 papers published from 2000 to 2018, presenting detailed information on the research's spatial and temporal aspects, crime data, and forecasting methodologies. It also provided multiple summaries, including the top four proposed methods, the best-proposed methods, and the baseline methods applied in the selected papers. The study discussed the strengths, weaknesses, threats, and opportunities of these papers and concluded that the spatial continuity of algorithms should not be overlooked in future research [2].

Table 1. Previous study related to the research topic

Ref.	**Problem area**	**Data Type**	**Methods**	**Outcome**	**Limitation**
[1]	Problem of predicting criminal activity	String	K-means clustering, Naïve Bayesian Classification, Simulation Environment	Predictive accuracy up to 70%	Does not use large dataset.
[2]	Importance of fairness in machine learning applications in policing	String	Machine Learning	A proper understanding of the fairness of predictive policing approach	Limited academic research
Ref	**Problem area**	**Data Type**	**Methods**	**Outcome**	**Limitation**
[3]	The study addresses the limitations of the NYPD	String	Random Forest classifier	Accuracy of 22%	This ML model can't handle complex rand

	current predictive accuracy				rapid changes in urban crime
[4]	Socio-environmental factors contributing to crime mapping and their spatio-temporal distribution.	String	Risk Terrain Modelling (RTM) and GAMLSS R package	Effectively measures crime risks by linking past incidents to socio-environment factors	Can’t work with real time data

CRIEM DATASETS

Crime-related data are collected from a wide array of sources, encompassing police reports, social media posts, news articles, and criminal records. The aggregation of such extensive data can be a challenging task [30]. These datasets can be found online in many countries or are obtained directly from police departments. In our research, we observed that the Chicago crime dataset is a popular choice for crime prediction systems. This is likely due to the city's large population and high crime rates, making it a valuable resource for studying and predicting criminal activity.

Table 2. Crime Type: Felony

Crime	Description
Murder/homicide	Non-negligent or intentional killing refers to the act of one person causing the death of another. This encompasses a range of situations, including suicides, fatalities resulting form negligence or accidents, international assaults leading to murder, and cases of justifiable homicides, which are typically categorized as aggravated assaults.
Burglary	Trespassing into a building with the intent to steal or commit and serious crime, or attempting to forcibly enter, in referred to as burglary.
Forcibly rape	Forcing a female, regardless of her age into a sexual assault against her will through physical coercion in termed as rape. This encompasses instances of sexual assaults where there is a use of force or threat, leading to non-consensual sexual acts.
Illegal drug selling	This involves illegal activities related to drugs, such as drug trafficking and drug distribution, which encompass selling, transporting, and distributing narcotics. These actions are classified as federal crimes and are considered felonies, carrying significant penalties user the law.

Robbery	This refers to the act of trying to take something valuable from someone by using force or threats, which instill fear in the victim, and it can involve taking an item from their possession, control, or care.
Aggravated assault, battery	An unlawful assault where one person attacks another, often using a weapon, resulting in the victim experiencing significant bodily harm or obvious severe injuries.
Arson	Maliciously setting fire to, or intentionally attempting to burn, a motor vehicle, dwelling house, public building, aircraft, or someone else's personal property, with or without the intent of default, is referred to as arson.
Forgery	Counterfeiting in the act of duplicating, mimicking, or modifying something without proper authorization or legal right, with the aim to deceive or defraud by presenting the altered item as genuine or original, often for the purpose of buying or selling while intending to deceive or commit fraud.

Table 3. Crime Type: Misdemeanor

Crime	**Description**
Larceny-theft	This refers to the unlawful act of removing or taking property from someone else's possession, whether by physically carrying it. Leading it away, riding it away, or any other means. Examples of such acts include stealing motor vehicle parts, bicycle theft, shoplifting and pickpocketing.
Fraud	A deliberate misrepresentation of the truth with the intention of convincing another individual or entity or relinquish a legal right or something valuable is known as fraud.
Embezzlement	This signifies the unlawful act of an individual diverting or appropriating for their own purpose's property, money, or another valuable item that has been entrusted to their care and custody.
Stolen property	It involves the actions of receiving, selling, buying, concealing, possessing, or transporting any property while being aware that it has been acquired through illegal means, such as fraud, theft, robbery, burglary, or embezzlement. These actions may also include attempted involvement is such activities.
Vandalism	This pertains to the deliberate act of damaging, altering, defacing, or harming any property, whether it's privately owned or public, personal or real estate, without the consent of the owner or the person in control or custody of it. This can include actions like tearing, marking, painting, cutting, breaking, drawing, covering with filth, or any other such means. The concept also encompasses attempted acts of this nature
Gambling	This involves the unlawful act of betting or wagering money on something of value, participating in, promoting, assisting, or operating betting activities, sharing betting information or transmitting it, as well as engaging in the acquisition, production, sale, or transportation of gambling equipment, devise, or goods.

Drunkenness	This refers to the act of consuming alcohol to an extent where one's mental abilities, faculties, and physical coordination are significantly compromised or impaired.

Table 4. Crime Type: Infraction and Wobblers

Crime	Description
Overtime parking	Staying in a designed parking area for a duration exceeding the time limit indicated.
Speeding ticket	It is a document issued by a police officer
Tailgating	This describes the hazardous and unlawful practice of driving near the vehicle in front. The driver of the leading vehicle suddenly applied the brakes. The tailgating driver faces a significant risk of a potential and unavoidable collision.
Weapons violation	This involves the act of holding, carrying, or engaging in actions that branch local ordinances or Laws prohibition the sale, concealment, transportation, possession, purchase, production, or use of sharp cutting tools, incendiary device, explosives, firearms, and related items.

3. Data and Methodology:

Various techniques and methodologies can be used to achieve the above-mentioned objective, the study and work region is Karachi city, which is widely spread, highly populated and having nested complex road network. In order to investigate the said goals in a relevant manner, actual & complete crime data is required and analyzed accordingly.

3.1. Data & Sources

For real world dataset, is said that it is never be as you expect. What you get is usually what you not require. Real datasets mostly unstructured, noisy, inconsistent incomplete and consisting missing data in records. It is mainly due to the lack of standardization, viable requirements and need of organizations. Thus, this trouble in data processing and knowledge discovery. This improper and imbalance dataset entail data cleaning to improve the data quality and make it more meaningful in information extraction. Data preprocessing is not an easy task which is actually consume around 70% of efforts of complete data analysis and predictive development [66].

Abating these issues, the dataset and other resources that are utilized that should render the factual ground knowledge of crimes in city. The expressive crime map presentation, analyzing and producing results requires the database/dataset which should at least contain the date, time, area, subarea, location base information of specific crime (GEO coordinates). In addition to this town-based information, UC based information, individual police station and its jurisdictional information and arrested or not. All these information is also necessary for better analysis, portraying crime pattern, criminal nature, groups, interlinks and their associative region. The possible attainment of such desirable data can be obtained by different sources such as CPLC, security agencies, government organization, and Karachi Police.

3.2. Data Collection

The hypothetical data is useful in understanding and structuring that made the system closest to portraying the real picture. The time, area, geo-location cords information are considered as the most important attributes for developing the system and mapping crimes. The google maps is used to collect the GEO coordinates with error approximate of 20 meters, as it plays and important role in location-based monitoring and mapping crimes.

3.3. Availability and Accuracy of Data

In predictive analysis, data is the key essential and important element of system. Most often the data required for modelling are obtained from databases or flat files which is full of errors. While, the most predictive models require clean formatted data[67], and its credibility is prerequisite condition of any such product. Whereas, the data selection is mainly based on the project nature and its ultimate goals to achieved[64].

The dataset used in project is of mobile crime contains the data from all over the Bangladesh, which has various attributes like case id, mobile brand, mobile model, crime type, area and landmark name, date and some un-useful and unnecessary attributes. While the provided dataset is not up to the mark as needed and missing the critical attributes like time, location code, police station jurisdiction information, town information and union council base information. In addition to data accuracy there are lots of multiple spelling error, improper and irrelevant area and landmark names are found.

3.4 Dataset Attributions

The given dataset of Karachi city is used for building the predictive model, analysis and tune system. The dataset consist of 16 attributes shown in Table II, and more than six lacs records from the period year 2005 to 2013.

We distinguish two basic roles a variable:

- Independent (also called predictor, explanatory, feature) – these variables describe the properties of objects which we want to use as the basis for making inferences.
- Dependent (also called response, explained, target) – these variables describe the features of the object which we want to make inferences about.

Table 5. *ORIGINAL ACQUIRED DATASET*

Attribute	**Type/Description**	**Attribute**	**Type/Description**
COMPNO	Int/Complain number	INCDATE	Dt/Incident date

COMPDATE	Dt/Complain date	AREANAME	Txt/Incident area name
IMEINUMB ER	Int/Mobile IMEI number	INCLANDMR K	Txt/Incident location name
SIMNUMBE R	Int/Mobile SIM number	PSNAME	Txt/Police station name
BRANDNA ME	Txt/Mobile brand name	COMPNAME	Txt/Complainer name
MODELNA ME	Int/Mobile model number	CNICNUMBE R	Int/Complainer CNIC number
COLNAME	Txt/Mobile color	CADDRESS	Txt/Complainer address
CRNAME	Txt/Crime type	CHOMEPHO NE	Txt/Complainer contact number

4. Data Pre-Processing

The pre-processing data comprises of several tasks which include cleaning of inconsistent and noise in information, data integrating process, transforming the dataset into intended and machine process-able form and save it into data warehouse or storage location[66]. Furthermore, in addition to this, adding, splitting, merging, and extracting necessary and hidden data attributes are also done in this step as shown. The resultant dataset is then feed into predictive model that is processed by the machine learning algorithm for training and testing[68], [69]. Basic exploratory observation found that the dataset have missing values and number of inoperable implausible attributes. Among the above mentioned attributes, this work only focusses on potential crime related fields. Because, the performance efficiency and quality of knowledge discover process in machine learning application is directly proportional and dependent on the pre-processed or fed data quality[66]. Following are a number steps taken to clean the dataset and illustrated and transformed dataset is shown

- Import appropriate data into environment.
- Explore dataset to determine data health.
- Select data range or records to process.
- Identify potential and unusable attributes.
- Replace and fill missing information of record(s) with default, mean, model derived or global constant value(s).
- Remove incomplete information record(s), if step 6 not fill the gap(s)
- Remove the complete record(s) contain(s) garbage data.
- Remove redundant and duplicate record(s) based on attributes tuples (INCDATE, COMPDATE, COMPNAME, IMEINUMBER). The given dataset is merge of various sources, this tuple helps in identifying duplicate or redundant entry in dataset.

- Remove unusable attribute(s): COMPNO, PSNAME, SIMNUMBER, IMEINUMBER, SIMNUMBER, MODELNAME, CNICNUMBER, COLNAME, CADDRESS, and CHOMEPHONE.
- Split attribute(s) into new potential attribute(s): INCDATE fragmented into day, month, year, weekday name i.e Tuesday, week number and reported time.
- Add appropriate or extracted attribute(s): Add Town and incident location geo-coordinates with help of Google API and named them as MALong, MALat for town longitude and latitude, SALong, SALat for incident location.
- Grouping attribute(s) values for more meaningful information: Time into Time four (4) categorical time slap window 12am-5:59am, 06am-11:59am, 12pm-5:59pm, 06pm-11:59pm.
- Removing multi-spell error by renaming values to standard spell as used by Google which helps in finding coordinate and also will be helpful to plot the final results on map.
- Save transformed dataset as clean Dataset.

Table 6. PROCESSED TRANSFORMED DATASET

Attributes	Description	Type
Brand	It shows Mobile Brands, which consist of 8 distinct mobile brands; Black Barry, HTC, iPhone, LG, Megagate, Motorola, Nokia, Samsung, and Sony are numbered as 1 to 9 in alphabetical order.	Numeric
Crime	Crime Type focusing on only two street crimes; Snatch and Theft	Numeric
Date	Incident Date. Calendar date	Date
Day	Incident day of week in number, where Monday is day 1.	Numeric
Week	Week of the year in number i.e. 27, 28, 29….	Numeric
Month	Denotes the month of the year, where January is 1 and December is 12	Numeric
Area	Incident area (main area) name	Text
Landmark	Incident sub-area (landmark, street, etc.) name	Text
MALong	Main area Latitude	Numeric
MALat	Main Area Longitude	Numeric
SALong	Sub-area Latitude	Numeric
SALat	Sub-area Longitude	Numeric

5. TECHNIQUE

Exploratory Data Analysis (EDA) is a heart of predictive system that was named by John Tukey, use for analyzing and summarizing dataset by their characteristics with visual techniques[65]. This approach does not necessary require any statistical model but can be used for knowing what data reveals without hypothetical and formal model testing[66]. EDA visual techniques are generally very simple in nature and quite expressive that easy to understand. These consist of:

Figure 1. Proposed predictive policing system architecture

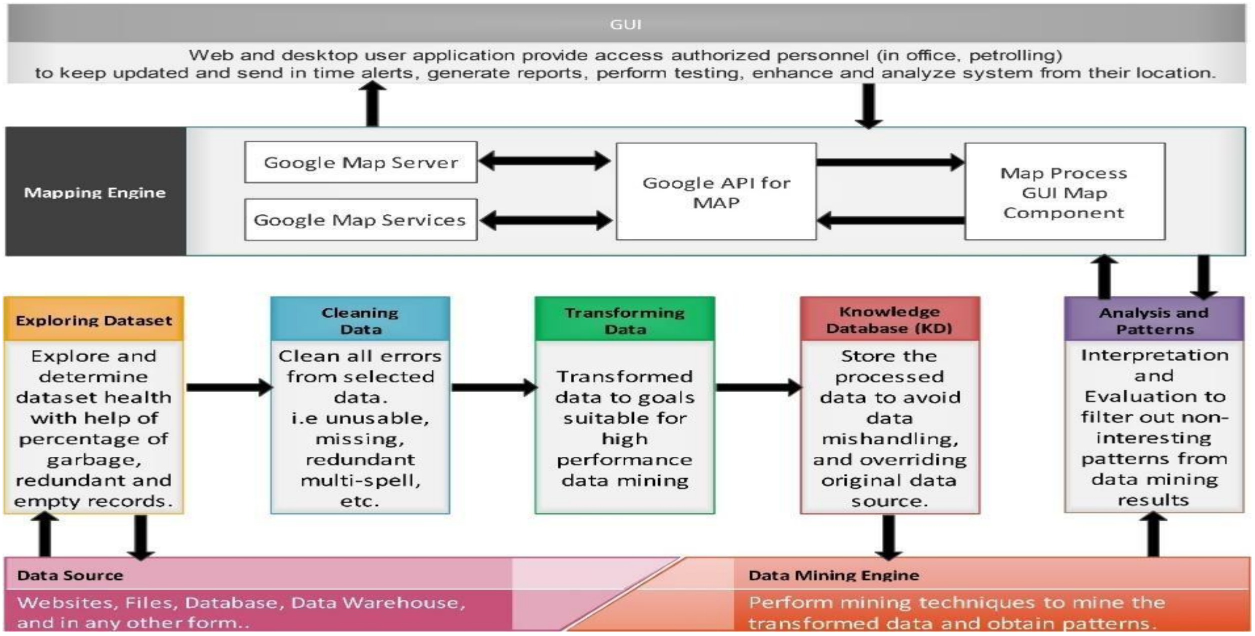

GUI
Web and desktop user application provide access authorized personnel (in office, petrolling) to keep updated and send in time alerts, generate reports, perform testing, enhance and analyze system from their location.
Mapping Engine
Google Map Server
Google Map Services
Google API for MAP
Map Process GUI Map Component
Exploring Dataset
Explore and determine dataset health with help of percentage of garbage, redundant and empty records.
Cleaning Data
Clean all errors from selected data. i.e unusable, missing, redundant multi-spell, etc.
Transforming Data
Transformed data to goals suitable for high performance data mining
Knowledge Database (KD)
Store the processed data to avoid data mishandling, and overriding original data source.
Analysis and Patterns
Interpretation and Evaluation to filter out non-interesting patterns from data mining results
Data Source
Websites, Files, Database, Data Warehouse, and in any other form..
Data Mining Engine
Perform mining techniques to mine the transformed data and obtain patterns.

Figure 2. Data cleaning process flow

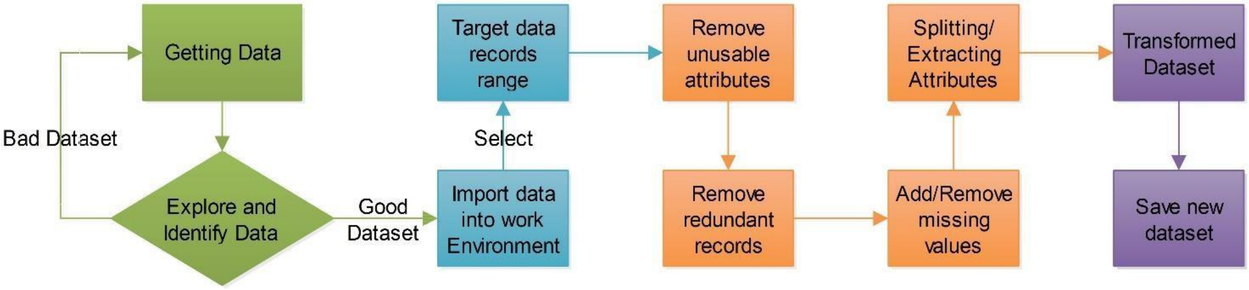
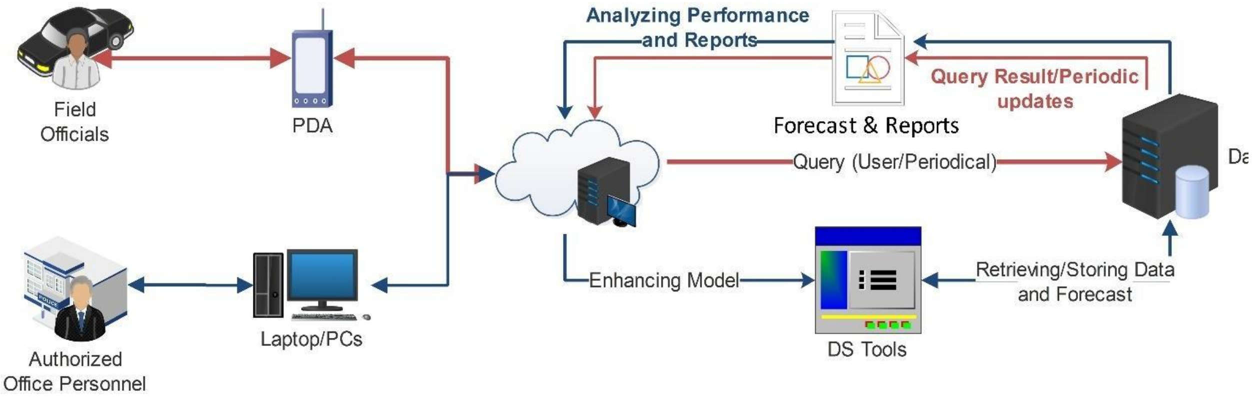

Getting Data
Bad Dataset
Explore and Identify Data
Good Dataset
Import data into work Environment
Select
Target data records range
Remove unusable attributes
Remove redundant records
Add/Remove missing values
Splitting/ Extracting Attributes
Transformed Dataset
Save new dataset

Figure 3. Predictive policing system process model


Field Officials
PDA
Authorized Office Personnel
Laptop/PCs
Analyzing Performance and Reports
Forecast & Reports
Query Result/Periodic updates
Query (User/Periodical)
Da
Enhancing Model
DS Tools
Retrieving/Storing Data and Forecast

- Raw data graphs such as histograms plot, box plot, and probability plots.

- Statistical graphs such as standard deviation, box, mean plots and others.
- Positioning graphs such as max, min, pattern- recognition, multiple plots and many more.

6. SYSTEM MODEL

The underlying system flow design serve clients to process, store and retrieve the data shown in Fig. 3. According to the present knowledge & research study, presented the design is fully optimized and flexible to for future enhancement and adjustment. New factors can be easily added to the system to process dynamically in order to improve the effectiveness and efficiency of system.

7. ALGORITHMS AND MINING TECHNIQUE

Since the idea enlighten the possibilities of making machines intelligent to use their advance and high computational power assist-able in better decision making, early alarming time sensitive applications. People have been proposed numerous machine learning algorithms that are designated to deal with their respective problem approach and broadly categorized them into two subject; Clustering and Classification.

7.1. K-means

Among the other machine learning algorithms, K- means[67], [34] is one of the simplest and less complex clustering algorithm[38] which comes under unsupervised learning technique[68]. Clustering mechanisms are mainly use for partitioning the data into their respective heads based on the characteristics similarities[31], [63], it does not predict future but sorting data into partitions. This helps in cluster analysis to learn the behaviour of corresponding entity to identify which geo-spatial region it belongs to[1]. In the following presents flavoured the K-means algorithm used in this work for sorting dataset.

- Setting number of K cluster as number of Main area in dataset. i.e 93.
- Sorting dataset w.r.t to main area, centroid is determine by the Longitude and Latitude of main area.
- Nearness Proximity is determined by Euclidean distance, cosine similarity, Haversine formula to estimate distance and also will help in mapping crime.
- Grouping the sub-area under their respective main areas.
- Repeats the iteration unless centroid stop changing
- End if all are stable and partitioned.
- Save the processed dataset.

7.2. Naïve Bayesian

Unlike K-means algorithm, Naïve Bayes[21] is one of the supervised and well-known classification machine learning approach use for predicting the future instances[22]. It has been extensively utilized in various studies and research works produced surprising result in innumerable domain[6] which is known for its better performance and accuracy as compare with others[52]. It uses Bayes theorem for computing the probability of every class from underlying evidence[1]. The general naïve Bayes equation used in building predictive model is shown in (1)

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times P(x_3|c) \times \ldots P(x_n|c) \times P(c)$$

$$\mathrm{P}_c = \frac{\mathrm{P}(crime|main\ area,\ sub\ area,\ \ date,\ day)}{P(main\ area, sub\ area, date, day)}$$

$$\begin{aligned} P_c = P(main\_area) \times P(cirme|main\_area) & \\ \times P(sub\_aera) & \\ \times P(crime|sub\_area) \times P(date) & \\ \times P(crime|date) \times P(day) & \\ \times P(crime|day) & \end{aligned}$$

Where P(c|x) donates the posterior probability, P(x|c) is likelihood, P(x) is predictor probability and P(c) is class probability.

8. Methodology

Exploratory Data Analysis (EDA) is a heart of predictive system that was named by John Tukey, use for analyzing and summarizing dataset by their characteristics with visual techniques[70]. This approach does not necessary require any statistical model but can be used for knowing what data reveals without hypothetical and formal model testing[68]. EDA visual techniques are generally very simple in nature and quite expressive that easy to understand. These consist of………

Working requires more time

9. Result and Discussion:

The primary results of the study that were obtained by clustering and forecasting city crimes using K-means and Naïve Bayesian algorithms are presented in this section. Fig. 4 shows

the density-based clustering of crimes. Density clustering identifies six significant areas among the 93 unique main areas that are similar in character and can be further divided into two types.

The lower clusters zone includes the areas of Defense, Korangi, and Sadar, while the top clusters zone includes Gulshan-e-Iqbal, Gulistan-e-Jouhar, and Nazmabad. However, if we divide the clusters according to their kind as indicated in the table, we discovered an interesting fact: the people who live in Gulshan-e-Iqbal, Gulistani-e-Jouhar, and Defense are upper middle class and upper class. However, Korangi, Sadar, and Nazmabad are the busiest business and industrial areas of the city that have high traffic and activity areas.

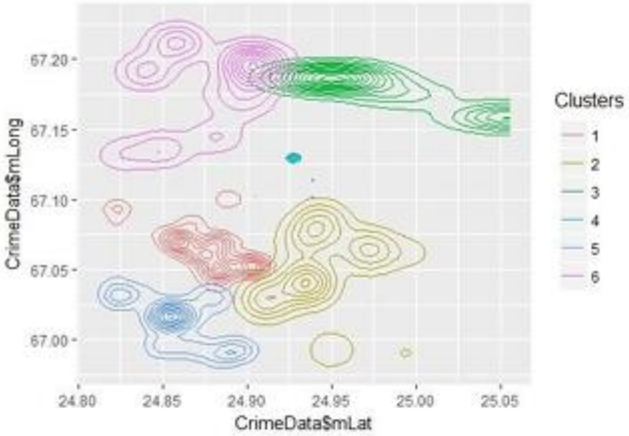
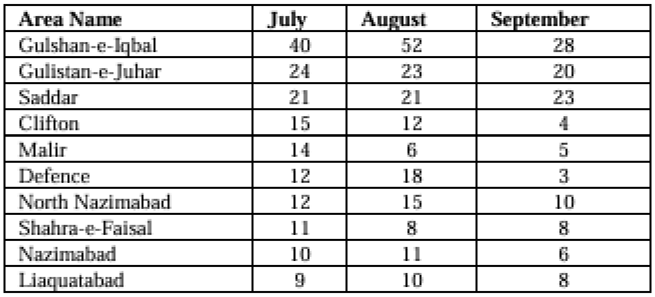


Figure 4. Crime density clusters

However, a few of the city's major economic and industrial districts, with lots of bustle and traffic, are Nazmabad, Sadar, and Korangi. Table IV lists the city's top ten crime hotspots out of 93. According to Table IV, the busiest parts of the city include Sadar, Gulshan-e-Iqbal, and Gulistan-e-Juhar, which also have the highest crime rates. Whereas the core trading market and headquarters of major corporations are in Sadar town, educational institutions and offices are primarily located in Gulshan-e-Iqbal and Gulistan-e-Juhar. Gulistan-e-Juhar and Gulshan-e-Iqbal are next to each other and have clear streets and broad roads, but Sadar Town is across the street from them and has narrow streets and heavy traffic. However, Gulistan-e-Juhar town, Sadar Twon, and Gulshan-e Iqbal are

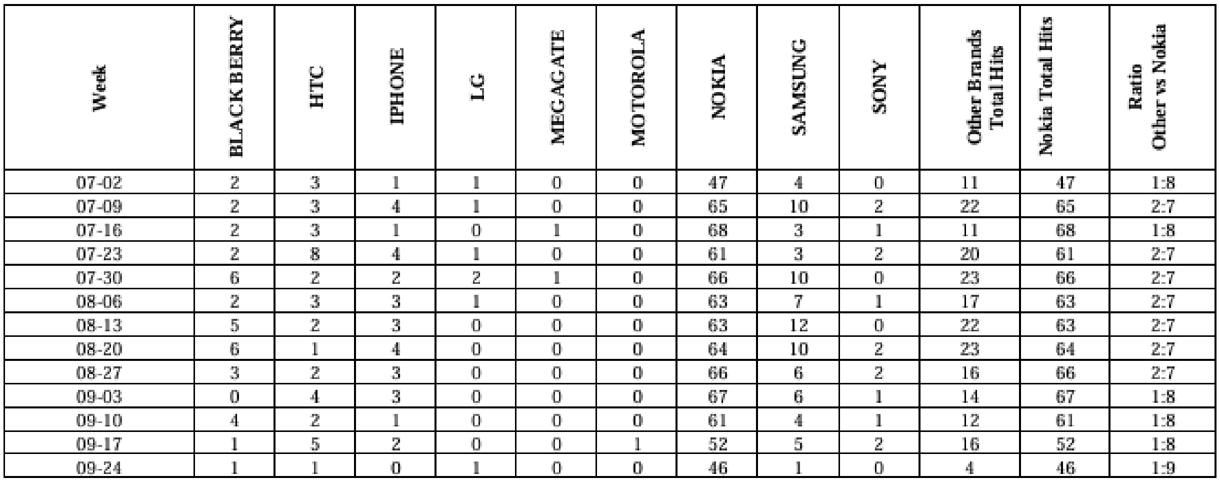
TABLE IV. TOP 10 CRIME AREAS



Area Name	July	August	September
Gulshan-e-Iqbal	40	52	28
Gulistan-e-Juhar	24	23	20
Saddar	21	21	23
Clifton	15	12	4
Malir	14	6	5
Defence	12	18	3
North Nazimabad	12	15	10
Shahra-e-Faisal	11	8	8
Nazimabad	10	11	6
Liaquatabad	9	10	8

The following Table V illustrates the weekly crime rate based on mobile brand. This reveals an interesting fact that among all brands Nokia is the only and highest lost rate or targeted brand. This also state that Nokia is the quit affordable and inexpensive brand among all or in other words based on city situation people more likely to buy the Nokia instead of any other brand which due to its affordability (low price), low loss cost, availability on every next mobile shop that makes it highly demanding mobile brand in the situation at that time. The probability of hitting other bands with compare to Nokia is about 1:8 (Other: Nokia). In every 8 hits there might be chance to get the high end mobile, or in another way, only one out of eight (1:8) people is prefer to use high-end or other mobile brands instead of Nokia

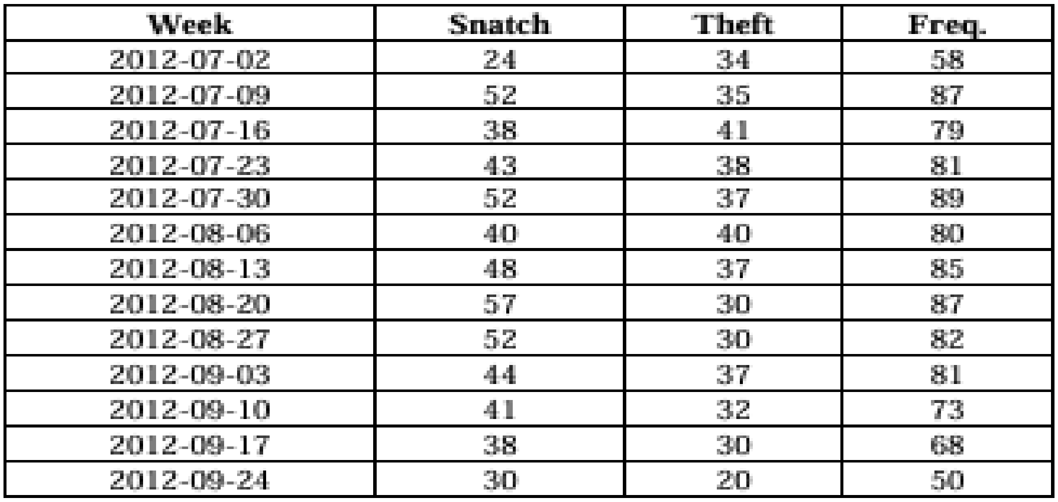
TABLE V. WEEKLY CRIME INTENSITY BY MOBILE BRAND



Week	BLACK BERRY	HTC	IPHONE	LG	MEGAGATE	MOTOROLA	NOKIA	SAMSUNG	SONY	Other Brands Total Hits	Nokia Total Hits	Ratio Other vs Nokia
07-02	2	3	1	1	0	0	47	4	0	11	47	1:8
07-09	2	3	4	1	0	0	65	10	2	22	65	2:7
07-16	2	3	1	0	1	0	68	3	1	11	68	1:8
07-23	2	8	4	1	0	0	61	3	2	20	61	2:7
07-30	6	2	2	2	1	0	66	10	0	23	66	2:7
08-06	2	3	3	1	0	0	63	7	1	17	63	2:7
08-13	5	2	3	0	0	0	63	12	0	22	63	2:7
08-20	6	1	4	0	0	0	64	10	2	23	64	2:7
08-27	3	2	3	0	0	0	66	6	2	16	66	2:7
09-03	0	4	3	0	0	0	67	6	1	14	67	1:8
09-10	4	2	1	0	0	0	61	4	1	12	61	1:8
09-17	1	5	2	0	0	1	52	5	2	16	52	1:8
09-24	1	1	0	1	0	0	46	1	0	4	46	1:9

Table VI presents the weekly crime rate and crime distribution by type i.e: Snatch and Theft. The statistics of Table 6 and recalling the above mentioned parameters of top 3 highly crime areas in our analysis i.e: Gulshan-e-Iqbal, Gulistan-e- Juhar town and Sadar Twon. We can conclude that it is comparatively more comfortable and easy for criminals to snatch your mobile or lost mobile while people are in high congested market area or walking alongside on wide running road, instead of someone silently pickup your pocket. From the Dataset, we found that the majority of theft attempts are happened in market areas i.e. Sadar town (is one of market areas) while snatching incidents are majorly reported in other areas which covers institutional areas, private company or offices and restaurants situated in commercial lane around residential areas i.e Gulshan-e-Iqbal, Gulistan-e- Juhar

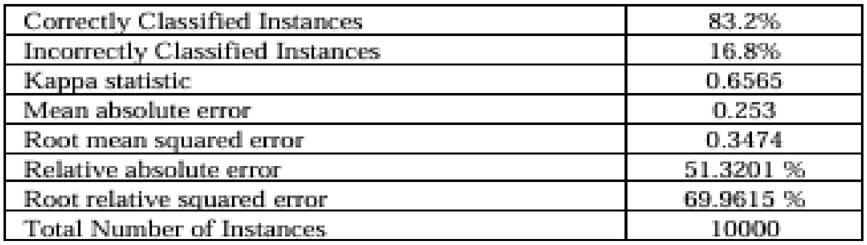
TABLE VI. WEEKLY CRIME RATE BY CRIME TYPE



Week	Snatch	Theft	Freq.
2012-07-02	24	34	58
2012-07-09	52	35	87
2012-07-16	38	41	79
2012-07-23	43	38	81
2012-07-30	52	37	89
2012-08-06	40	40	80
2012-08-13	48	37	85
2012-08-20	57	30	87
2012-08-27	52	30	82
2012-09-03	44	37	81
2012-09-10	41	32	73
2012-09-17	38	30	68
2012-09-24	30	20	50

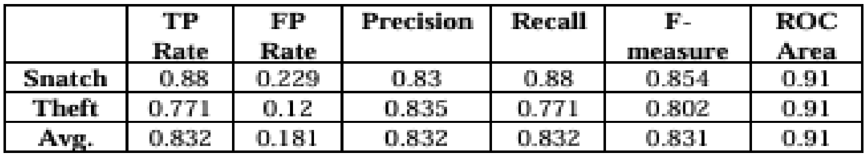
Table VII shows predictive model performance summary. Model is trained on different scale ratio in different cycle to test and identify the most suitable point to splitting dataset with the window of 5% from 60% to 90% dataset, and the remaining was used for testing. Experiments found that the optimum result of model using NB gained at 80-20 ratio. Above the 84% or below 80% the accuracy of model gradually decreasing and the min 67.8% accuracy recorded. Minimum 5 related features are require to find out the expected next hit. These potential features are month, week, day, time, and crime type. Table VIII shows the prediction accuracy of model in which we are identifying what type of crime could be happen in future. We successfully achieved significant accuracy of about 83%.

TABLE VII. PREDICTION SUMMARY



Correctly Classified Instances	83.2%
Incorrectly Classified Instances	16.8%
Kappa statistic	0.6565
Mean absolute error	0.253
Root mean squared error	0.3474
Relative absolute error	51.3201 %
Root relative squared error	69.9615 %
Total Number of Instances	10000

TABLE VIII. PREDICTION ACCURACY BY CLASS



	TP Rate	FP Rate	Precision	Recall	F-measure	ROC Area
Snatch	0.88	0.229	0.83	0.88	0.854	0.91
Theft	0.771	0.12	0.835	0.771	0.802	0.91
Avg.	0.832	0.181	0.832	0.832	0.831	0.91

Based on the given dataset (July, 12 – Sep-12), Fig. 5 depicts the next day hit. It is clearly seen that potential crime scene would be Gulshan-e-Iqbal, University road, and the Gulistan e-Jouhar area are under threat. From the cluster analysis we know that these areas have more crime density and other crime instances (near Frere and Nasir colony) which has very low probability of occurrence.

10. Conclusion:

Inspired by current criminology research, the opportunity for future projection through advanced technology, frequent occurrences of terrible incidents, uncertainty, and ensuring human life safety and improving the strategies used by law enforcement. The significance, benefits, and widespread use of predictive policing by multiple countries that demonstrate positive outcomes quickly are discussed in this paper. To detect and deter future criminal activity, this study presented the Crime Analyst and Predictor (CAP). The first suggested system is based on the most advanced machine learning techniques; less complex methods like K-means clustering and Naïve Bayesian classification do better than others. The unstructured and noisy Karachi Street crime dataset, which contains records of mobile snatching and theft, is the subject of this paper. First, the data was cleaned, converted into new data, and then fed as input into the constructed model. The model first used K-means to divide the data into clusters, and then it used naïve Bayesian to predict the crime date and place as an output. The simulation is run in the R and Weka environments, and the results indicate an accuracy of about 70%, which is very positive and encouraging for the ultimate real-time predictive application. The suggested system will be developed in future work in collaboration with the local government authorities and the industrial partner O'Reference liaison to transform the concept into a useful real-time application that maximizes the effective scare resource allocation.

Reference:

1. Khan, Jibran Rasheed, et al. "Predictive policing: A machine learning approach to predict and control crimes in metropolitan cities." University of Sindh Journal of Information and Communication Technology 3.1 (2019): 17-26.
2. Alikhademi, Kiana, et al. "A review of predictive policing from the perspective of fairness." Artificial Intelligence and Law (2022): 1-17.
3. *Safat W, Asghar S, Gillani SA. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. IEEE Access J. 2021;9:70080–94.*
4. *Kounadi O, Ristea A, Araujo A, Leitner M. A systematic review on spatial crime forecasting. Crime Sci. 2020;9(1):1–22.*
5. *Tollenaar N, van der Heijden PGM. Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. J R Stat Soc Ser A. 2013;176(2):565–84.*

6. *Enzmann D, Podana Z. Official crime statistics and survey data: Comparing trends of youth violence between 2000 and 2006 in cities of the Czech Republic, Germany, Poland, Russia, and Slovenia. Eur J Crim Policy Res. 2010;16(3):191–205.*
7. *Holst A, Bjurling B. A Bayesian parametric statistical anomaly detection method for finding trends and patterns in criminal behavior. In 2013 European Intelligence and Security Informatics Conference. IEEE; 2013.*
8. *Brunsdon C, Corcoran J, Higgs G. Visualising space and time in crime patterns: A comparison of methods. Comput Environ Urban Syst. 2007;31(1):52–75.*
9. *Vural MS, Gök M, Yetgin Z. Generating incident-level artificial data using GIS based crime simulation. In 2013 International Conference on Electronics, Computer and Computation (ICECCO). IEEE; 2013.*
10. *Xiang Y, Chau M, Atabakhsh H, Chen H. Visualizing criminal relationships: Comparison of a hyperbolic tree and a hierarchical list. Decis Support Syst. 2005;41(1):69–83.*
11. *Jain LC, Seera M, Lim CP, Balasubramaniam P. A review of online learning in supervised neural networks. Neural Comput Appl. 2014;25(3):491–509.*
12. *EL Aissaoui O, EL Madani EY, Oughdir L, EL Allioui Y. Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Procedia Comput Sci. 2019;148:87–96.*
13. *Mackenzie DM. CDUL: class directed unsupervised learning. Neural Comput Appl. 1995;3(1):2–16. 12. Rossmo DK, Laverty I, Moore B. Geographic profiling for serial crime investigation, in Geographic information systems and crime analysis. IGI Glob. 2005;6:102–17.*
14. *Ristea A, Leitner M. Urban crime mapping and analysis using GIS. ISPRS Int J Geo-Information. 2020;9(9):511.*
15. *Corcoran JJ, Wilson ID, Ware JA. Predicting the geotemporal variations of crime and disorder. Int J Forecast. 2003;19(4):623–34.*
16. *Sangani A, Sampat C, Pinjarkar V. Crime prediction and analysis. In 2nd International Conference on Advances in Science & Technology (ICAST); 2019.*
17. *Wang Y, Peng X, Bian J. Computer crime forensics based on improved decision tree algorithm. J Netw. 2014;9(4):1005.*
18. *Khan M, Ali A, Alharbi Y. Predicting and preventing crime: A crime prediction model using San Francisco crime data by classification techniques. New Jersey: Wiley/Hindawi. Vol. 2022, No. 4830411, 2022. p. 13.*
19. *Ewart BW, Oatley GC. Applying the concept of revictimization: using burglars' behaviour to predict houses at risk of future victimization. Int J Police Sci Manag.2003;5(2):69–84.*
20. *Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time seriesanalysis: forecasting and control. John Wiley & Sons; 2015.*
21. *Jangra M, Kalsi S. Naïve Bayes approach for the crime prediction in Data Mining. Int J Comput Appl. 2019;178(4):33–7.*
22. *Khairuddin A, Alwee R, Haron H. A comparative analysis of artificial intelligence techniques in forecasting violent crime rate. In IOP Conference Series: Materials Science and Engineering. IOP Publishing; 2020.*
23. *Sardana D, Marwaha S, Bhatnagar R. Supervised and unsupervised machine learning methodologies for crime pattern analysis. Int J Artif Intell Appl. 2021;12(1):43–58.*
24. *Sivanagaleela B, Rajesh S. Crime analysis and prediction using fuzzy c-means algorithm. In 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE; 2019.*
25. *Liu X, Sun H, Han S, Han S, Niu S, Qin W, et al. A data mining research on office building energy pattern based on timeseries energy consumption data. Energy Build. 2022;259:111888*
26. *Borowik G, Wawrzyniak ZM, Cichosz P. Time series analysis for crime forecasting. In 2018 26th International Conference on Systems Engineering (ICSEng). IEEE; 2018.*
27. *Goel A, Singh B. White collar crimes: A study in the context of classification, causation and preventive measures. Contemp Soc Sci. 2018;27:84–92.*
28. *Grove L, Farrell G. Once bitten, twice shy: Repeat victimization and its prevention. Oxf Handb Crime Prev. 2012;404–19.*
29. *Chainey SP, da Silva BFA. Examining the extent of repeat and near repeat victimisation of domestic burglaries in Belo Horizonte, Brazil. Crime Sci. 2016;5(1):1–10.*
30. *Peter P, Ickjai L. Crime analysis through spatial areal aggregated density patterns. Geoinformatica. 2011;15(1):49–74.*
31. *30. Jonas p, Paul E, Stijn V, Marc MVH, Guido D. Gaining insight in domestic violence with emergent self organizing maps. Expert Syst Appl. 2009;36(9):11864–74.*

32. *Kang HW, Kang HB. Prediction of crime occurrence from multi-modal data using deep learning. PLoS One. 2017;12(4):e0176244.*
33. *Shekhar S, Evans MR, Kang JM, Mohan P. Identifying patterns in spatial information: A survey of methods. Wiley Interdiscip Reviews Data Min Knowl Discovery. 2011;1(3):193–214.*
34. *Mookiah L, Eberle W, Siraj A. Survey of crime analysis and prediction. In The Twenty-Eighth International Flairs Conference; 2015.*
35. *Hassani H, Huang X, Silva ES, Ghodsi M. A review of data mining applications in crime. Stat Anal Data Mining: ASA Data Sci J. 2016;9(3):139–54.*
36. *Falade A, Azeta A, Oni A, Odun-ayo I. Systematic literature review of crime prediction and data mining. Rev Comput Eng Stud. 2019;6(3):56–63.*
37. *Okeke OC. An overview of crime analysis, prevention and predicton using data mining based on real time and location data. Int J Recent Technol Eng. 2022;5(10):99–103.*
38. *Kianmehr K, Alhajj R. Crime hot-spots prediction using support vector machine. In IEEE International Conference on Computer Systems and Applications. IEEE Computer Society; 2006.*
39. *Antolos D, Liu D, Ludu A, Vincenzi D. Burglary crime analysis using logistic regression. In International Conference on Human Interface and the Management of Information. Berlin: Springer; 2013.*
40. *Cavadas B, Branco P, Pereira S. Crime prediction using regression and resources optimization. In Portuguese Conference on Artificial Intelligence. Springer; 2015.*
41. *Cesario E, Catlett C, Talia D. Forecasting crimes using autoregressive models. in 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE; 2016.*
42. *Vural MS, Gök M. Criminal prediction using Naive Bayes theory. Neural Comput Appl. 2017;28(9):2581–92.*
43. *Hou M, Hu X, Cai J, Han X, Yuan S. An integrated graph model for spatial–temporal urban crime prediction based on attention mechanism. ISPRS Int J Geo-Information. 2022;11(5):294.*
44. *Ilhan F, Tekin SF, Aksoy B. Spatio-temporal crime prediction with temporally hierarchical convolutional neural networks. In 2020 28th Signal Processing and Communications Applications Conference (SIU). IEEE; 2020.*
45. *Meskela TE, Afework YK, Ayele NA, Teferi MW, Mengist TB. Designing time series crime prediction model using long short term memory recurrent neural network. Int J Recent Technol Eng. 2020;9:402–5.*
46. *Hussain FS, Aljuboori AF. A crime data analysis of prediction based on classification approaches. Baghdad Sci J. 2022;4:1073–7.*
47. *Lin YL, Yen MF, Yu LC. Grid-based crime prediction using geographical features. ISPRS Int J Geo-Information. 2018;7(8):298.*
48. *Stec A, Klabjan D. Forecasting crime with deep learning. arXiv preprint arXiv; 2018. p. 01486.*
49. *Kim KS, Jeong YH. A study on crime prediction to reduce crime rate based on artificial intelligence. Korea J Artif Intell. 2021;9(1):15–20.*
50. *Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F, Pentland A. Once upon a crime: towards crime prediction from demographics and mobile data. In Proceedings of the 16th International Conference on Multimodal Interaction; 2014.*
51. *Zhuang Y, Almeida M, Morabito M Ding W. Crime hot spot forecasting: A recurrent model with spatial and temporal information. In 2017 IEEE International Conference on Big Knowledge (ICBK). IEEE; 2017.*
52. *Ivan N, Ahishakiye E, Omulo EO, Taremwa D. Crime prediction using decision tree (J48) classification algorithm. International Journal of Computer and Information Technology. 2017;6:188–95.*
53. *El Bour HA, Ounacer S, Elghomari Y, Jihal H, Azzouazi M. A crime prediction model based on spatial and temporal data. Periodicals Eng Nat Sci. 2018;6(2):360–4.*
54. *Kim S, Joshi P, Kalsi PS, Taheri P. Crime analysis through machine learning. In IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE; 2018.*
55. *Bharati A, RA KS. Crime prediction and analysis using machine learning. Int Res J Eng Technol (IRJET). 2018;5:1037–42.*
56. *Mahmud S, Nuha M, Sattar A. Crime Rate Prediction Using Machine Learning and Data Mining, in Soft Computing Techniques and Applications. Singapore: Springer; 2021. p. 59–69.*
57. *Almuhanna AA, Alrehili MM, Alsubhi SH, Syed L. Prediction of crime in neighbourhoods of New York City using spatial data analysis. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA). IEEE; 2021.*

58. *Yu CH, Ward MW, Morabito M, Ding W. Crime forecasting using data mining techniques. In 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE; 2011.*
59. *Sharma H, Kumar S. A survey on decision tree algorithms of classification in data mining. Int J Sci Res. 2016;5(4):2094– 7.*
60. *Gupta A, Mohammad A, Syed A, Halgamuge MN. A comparative study of classification algorithms using data mining: crime and accidents in Denver City the USA. Education. 2016;7(7):374–81.*
61. *Boppuru PR, Ramesha K. Spatio-temporal crime analysis using KDE and ARIMA models in the Indian context. Int J Digital Crime Forensics. 2020;12(4):1–19.*
62. *Tayal D, Jain A, Arora S, Agarwal S, Gupta T, Tyagi N. Crime detection and criminal identification in India using data mining techniques. AI Soc. 2015;30(1):117–27.*
63. *Iqbal R, Murad MAA, Mustapha A, Panahy PHS, Khanahmadliravi N. An experimental study of classification algorithms for crime prediction. Indian J Sci Technol. 2013;6(3):4219–25.*
64. *Almanie T, Mirza R, Lor E. Crime prediction based on crime types and using spatial and temporal criminal hotspots. arXiv preprint arXiv; 2015. p. 02050.*
65. *Yerpude P, Gudur V. Predictive modelling of crime dataset using data mining. Int J Data Min Knowl Manag Process. 2020;7:83–99.*
66. *Prathap BR, Krishna A, Balachandran K. Crime analysis and forecasting on spatio temporal news feed data—An indian context, in artificial intelligence and blockchain for future cybersecurity applications. Switzerland: Springer; 2021. p. 307–27*