RESEARCH
O&G Forum 2024; 34 – 2s: 25-32

UTILISING MACHINE LEARNING TO FORECAST AND ANALYSE CRIME RATES

Vedatrayee Chatterjee[1], Arnab Maitra[2], Hritik Banerjee[3], Subhadeep Puitandi[4], Soubhik Ghosh[5], Ankita Mukherjee[6]
[1]Assistant Professor, Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. vedatrayeechatterjee@gmail.com
[2] Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. arnabmaitra11@gmail.com
[3] Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. banerjeehritik@gmail.com
[4] Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. subhadeep8250@gmail.com
[5] Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. souvikghosh922@gmail.com
[6] Department of Computer Science & Engineering, Asansol Engineering College, Asansol, India. ankitamukherjee873@gmail.com

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.
Keyword: 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.

CRIME 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 1A. 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 from negligence or accidents, intentional 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 a serious crime, or attempting to forcibly enter, is referred to as burglary.
Forcibly rape	Forcing a female, regardless of her age, into a sexual assault against her will through physical coercion is 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 under 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 to defraud, is referred to as arson.
Forgery	Counterfeiting is 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 1B. 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 to 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 purposes 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 in 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, devices, 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 1C. Crime Type: Infraction and Wobblers

Crime	Description
Overtime parking	Staying in a designated parking area for a duration exceeding the time limit indicated.
Speeding ticket	It is a document issued by a police officer to a driver who has exceeded the speed limit, indicating that the driver is required to pay a fine.
Tailgating	This describes the hazardous and unlawful practice of driving in close proximity to the vehicle in front. If the driver of the leading vehicle were to suddenly apply 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 breach local ordinances or laws prohibiting the sale, concealment, transportation, possession, purchase, production, or use of sharp cutting tools, incendiary devices, explosives, firearms, and related items.

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 back 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 in size, 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].

CLASSIFICATION OF PREDICTION SYSTEMS

In the realm of machine learning, classification within prediction systems involves creating a model that can elucidate the categories or classes of data. The primary goal of this model is to predict the class of objects that lack a known class label. In the real world, law enforcement agencies grapple with the challenge of managing a vast number of criminal activities. Given the escalating crime rates, there is an imperative to leverage data mining in conjunction with police efforts to predict and subsequently mitigate these criminal cases. As emphasized, technology, particularly computer science tools, is indispensable in expeditiously addressing this issue. Prediction systems can be categorized based on various factors [36]:

a. In terms of methodologies, they can be classified into machine learning and data mining approaches.
b. Concerning the type of prediction, they may fall into spatial or temporal prediction categories.
c. Regarding the dataset, prediction systems can be further divided into image prediction and data prediction.

COMPARISON STUDY: CRIME PREDICTION VS CLASSIFICATION APPROACHES

This section presents Tables 2 and 3, which compile literature surveys of machine learning and data mining algorithms employed with various datasets across different cities globally. Furthermore, it conducts a comparative analysis between machine learning and data mining techniques concerning their application in border crime prediction systems. These tables provide a comprehensive overview of each selected paper, offering key details to aid fellow researchers in identifying the most effective categories of crime prediction techniques. In essence, these two tables elucidate the utilization of machine learning and data mining algorithms in the context of crime prediction, aligning with the objectives of this survey.

These tables include references to the selected papers, the machine learning or data mining algorithms employed, the sources of the datasets used, and the accuracy achieved by each algorithm with respect to specific datasets for various cities. The subsequent section delves into a discussion of research works in crime prediction, segregating them into those adopting machine learning approaches and those leveraging data mining methods.

A. Machine learning and crime prediction

The field of crime prediction has received extensive attention due to its profound societal implications. Machine learning algorithms have been instrumental in addressing crime prediction and forecasting challenges. These algorithms have proven successful in predicting spatial crime patterns. In 2006, the Support Vector Machine (SVM) algorithm was employed to predict crime locations in Columbus, Ohio, USA. SVM utilized both random and clustering approaches for training and testing datasets, thereby predicting hotspots and enhancing effectiveness [37]. These algorithms have been instrumental in exploring the link between crime occurrences and motivating factors. For instance, in 2013, a Logistic Regression (LR) algorithm was used to forecast the relationship between burglary crimes and various factors, such as time of day, day of the week, barriers, connectors, and repeated victimization. However, this model proved less effective for larger geographic areas [38].

Table 2. Survey on Crime Prediction Research Works with Machine Learning

Year	Method	Dataset	Classification technique	Acc. %
2014	Machine learning	London mobile and crime data	RF	70
2017	Machine learning	Portland data	RNN	75
			LSTM	81
			GRU	81
2017	Machine learning	UCI machine learning repository website	DT (J48)	94.25
2017	Machine learning	Chicago, Illinois (image dataset)	SVM	67.01
			KDE	66.33
			DNN	84.25
2018	Machine learning	Chicago crime data	DT	38
			RF	59
			Neural network	81
2018	Machine learning	Vancouver police department	KNN	39.9
			Boosted decision	43.2
2018	Machine learning	Chicago crime data	RNN	74.1
		Portland crime data		63.8
		Chicago crime data	CNN	72.7
		Portland crime data		62.9
2018	Machine learning	Taoyuan/Taiwan	DNN	83
			KNN	87
			SVM	88
2018	Machine learning	Chicago crime data	KNN	78
			NB	64
			DT	78
2021	Machine learning	Chicago crime data	LR	90
			DT	66
			RF	77
			MLP	87
			NB	73
			SVM	66
			XGBoost	94
			KNN	88

Year	Method	Dataset	Classification technique	Acc. %
		Los Angeles crime data	LR	48
			DT	60
			RF	43
			MLP	84
			NB	71
			SVM	60
			XGBoost	88
			KNN	89
2021	Machine learning	Bangladesh crime data	LR	73.6
2021	Machine learning	Baltimore city	Logistic regression	95
			Neural network	94
2021	Machine learning	New York crime data	SVM	43
			RF	50
			XGBoost	52

In 2015, crime prediction in southern US states was conducted using the Random Forest (RF) method. This approach was optimized by applying the SmoteR algorithm to identify more dangerous crimes. The study was further enhanced using R software, with density and population as real value selections [39]. Furthermore, the auto-regressive approach was adopted to predict and forecast the number of crimes occurring simultaneously in urban areas [40]. In 2017, the Naive Bayes (NB) algorithm was proposed to predict crime incidents based on historical data reflecting previous crimes in the same location. Additionally, the NB model was compared with the Decision Tree (DT) algorithm to assess performance, with NB outperforming DT despite DT's computational complexity [41].

In 2020, numerous research efforts introduced innovative methodologies. One study fused three methods - Long short-term memory (LSTM), Residual neural network, and Graph convolutional network - to extract spatial-temporal features for predicting crimes in Chicago. Evaluation metrics like Root mean square error and Mean absolute error were used to assess the method's performance [42]. Another study introduced a crime network for spatiotemporal data, utilizing Convolutional neural network (CNN) for the automated prediction of crime time and location [43]. Additionally, a Recurrent neural network (RNN) with LSTM was integrated to design a time series crime prediction system for Addis Ababa [44]. In yet another study, the severity level of crime in Boston was studied and predicted using machine learning algorithms such as SVM, NB, LR, and DT.
These machine learning endeavors have significantly contributed to understanding how crime behavior evolves over time. Several conclusions emerged from the survey presented in Table 2:

a. The choice between Deep neural network (DNN) and SVM depends on the dataset type, with DNN being more suitable for image crime datasets and SVM for text datasets.
b. The dataset significantly influences accuracy, as employing the same system on different crime datasets can yield notably different results.
c. Among machine learning approaches, the highest accuracy in crime prediction was achieved.
d. Logistic Regression (LR) demonstrated the highest accuracy among various machine learning algorithms.
e. When employing the RF method, the achieved accuracy was relatively low compared to other methods.
f. Standard deviations of crime prediction accuracies showed that SVM outperformed LR, achieving an accuracy of 71.9% and surpassing all other machine learning algorithms in terms of standard deviation results.

The highest accuracy in crime prediction results was often attained through the machine learning logistic regression method. For instance, it reached 95% for Baltimore city [48]. Algorithms such as XGBoost and Logistic Regression also demonstrated high accuracy levels of 94% and 90%, respectively [1]. However, it's essential to note that the same algorithm may perform differently with distinct datasets, underscoring the dataset's significant influence on crime prediction outcomes.

B. Data mining and crime prediction
In 2011, specialized data mining and technology techniques were introduced to extract patterns from spatial and temporal data. These methods involved geospatial analysis and were applied to predict residential burglaries in Portland using a dataset that included both spatial and temporal information. Various algorithms, including Naive Bayes (NB), Support Vector Machines (SVM), Decision Trees (DT), and K-Nearest Neighbors, were employed to forecast crimes. The results were compared, highlighting the effectiveness of neural networks in handling complex systems [57]. However, the usefulness of pattern extraction was somewhat limited due to the intricate relationships within the spatial data [32].
In 2016, high accuracy was achieved through the utilization of different Decision Tree algorithms for extracting insights from a dataset with 1994 instances and 128 attributes. A comparison was made between these algorithms, and the data were visualized using scatter plots to identify areas with varying levels of crime severity based on historical data [58].

Table 3. Survey on Crime Prediction Research Works with Data Mining

TABLE 3 - SURVEY ON CRIME PREDICTION RESEARCH WORKS WITH DATA MINING

Year	Method	Dataset	Classification technique	Acc.%
2013	Data mining	Different states of USA	DT	83.9
2014	Data mining	India crime data	WEKA on two K-mean clusters	93.62 for C1
				93.99 for C2
2015	Data mining	Denver crime data	NB	51
			DT	42
		Los Angeles crime data	NB	54
			DT	43
2017	Data mining	Chicago crime data	DT	75.9
			RF	83.39
			NB	77.64
2021	Data mining	Bangladesh crime data	NB	69.5
			KNN	76.9
2021	Data mining	Indian and Bangalore crimes data	KDE	77.49

In the same year, data mining algorithms were developed to classify crimes based on their types and time-related factors, such as academic schedules. These classifiers were used to predict the risk of crime severity in Denver between 2010 and 2015 [59].
In 2020, the Autoregressive Integrated Moving Average (ARIMA) technique was implemented to predict time series data, and the results were visualized using data mining tools. This technique demonstrated that regression models could effectively predict future crimes using historical newsfeed data [60].

Several important conclusions emerged from this survey:

a. Table 3 provides a comparative analysis of various algorithms used to address the challenge of crime prediction. These algorithms, such as Decision Trees (DT), Naive Bayes (NB), Random Forest (RF), and others, were applied individually or in combination to specific datasets and cities. This offered a significant challenge to these algorithms to verify their effectiveness and accuracy in crime prediction.
b. The survey in Table 3 highlights the algorithms that achieved the highest accuracy in crime prediction.
c. According to reference [61], the K-means algorithm demonstrated the highest accuracy among the various data mining algorithms.
d. Analyzing the standard deviations of crime prediction accuracies for each algorithm revealed that the Decision Tree (DT) algorithm outperformed the Naive Bayes (NB) algorithm, achieving an accuracy of 18.9%.

In previous studies, Decision Trees and Neural Networks recorded an impressive 94% accuracy for different datasets in references [48] and [51], showcasing their capabilities in machine learning algorithms. Furthermore, the K-means data mining algorithm achieved accuracy rates of 93.62% for cluster one and 93.99% for cluster two when applied to crime data in India [61].

CONCLUSION

The results indicate that data mining methods have achieved higher accuracy in crime prediction compared to machine learning methods. However, on average, machine learning tends to outperform data mining in this task. When we examine the standard deviation of crime prediction accuracies for both machine learning and data mining, it becomes evident that machine learning algorithms exhibit more consistency and perform better than data mining algorithms.

In summary, the comparison of machine learning and data mining algorithms for crime prediction systems suggests that the choice of an algorithm may depend on the specific dataset type (e.g., image, text, video, or voice dataset). It's also worth noting that certain algorithms excel on average but may not work as effectively with other types of datasets. It's important to mention that this survey did not cover crime prediction methods involving deep learning algorithms due to time limitations.

References

1. *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.*
2. *Kounadi O, Ristea A, Araujo A, Leitner M. A systematic review on spatial crime forecasting. Crime Sci. 2020;9(1):1–22.*
3. *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.*
4. *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.*
5. *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.*
6. *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.*
7. *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.*
8. *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.*
9. *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.*
10. *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.*
11. *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.*
13. *Ristea A, Leitner M. Urban crime mapping and analysis using GIS. ISPRS Int J Geo-Information. 2020;9(9):511.*
14. *Corcoran JJ, Wilson ID, Ware JA. Predicting the geo-temporal variations of crime and disorder. Int J Forecast. 2003;19(4):623–34.*
15. *Sangani A, Sampat C, Pinjarkar V. Crime prediction and analysis. In 2nd International Conference on Advances in Science & Technology (ICAST); 2019.*
16. *Wang Y, Peng X, Bian J. Computer crime forensics based on improved decision tree algorithm. J Netw. 2014;9(4):1005.*
17. *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.*
18. *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.*
19. *Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.*
20. *Jangra M, Kalsi S. Naïve Bayes approach for the crime prediction in Data Mining. Int J Comput Appl. 2019;178(4):33–7.*
21. *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.*
22. *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.*
23. *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.*
24. *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*
25. *Borowik G, Wawrzyniak ZM, Cichosz P. Time series analysis for crime forecasting. In 2018 26th International Conference on Systems Engineering (ICSEng). IEEE; 2018.*

26. *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.*
27. *Grove L, Farrell G. Once bitten, twice shy: Repeat victimization and its prevention. Oxf Handb Crime Prev. 2012;404–19.*
28. *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.*
29. *Peter P, Ickjai L. Crime analysis through spatial areal aggregated density patterns. Geoinformatica. 2011;15(1):49–74.*
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.*
31. *Kang HW, Kang HB. Prediction of crime occurrence from multi-modal data using deep learning. PLoS One. 2017;12(4):e0176244.*
32. *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.*
33. *Mookiah L, Eberle W, Siraj A. Survey of crime analysis and prediction. In The Twenty-Eighth International Flairs Conference; 2015.*
34. *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.*
35. *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.*
36. *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.*
37. *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.*
38. *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.*
39. *Cavadas B, Branco P, Pereira S. Crime prediction using regression and resources optimization. In Portuguese Conference on Artificial Intelligence. Springer; 2015.*
40. *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.*
41. *Vural MS, Gök M. Criminal prediction using Naive Bayes theory. Neural Comput Appl. 2017;28(9):2581–92.*
42. *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.*
43. *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.*
44. *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.*
45. *Hussain FS, Aljuboori AF. A crime data analysis of prediction based on classification approaches. Baghdad Sci J. 2022;4:1073–7.*
46. *Lin YL, Yen MF, Yu LC. Grid-based crime prediction using geographical features. ISPRS Int J Geo-Information. 2018;7(8):298.*
47. *Stec A, Klabjan D. Forecasting crime with deep learning. arXiv preprint arXiv; 2018. p. 01486.*
48. *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.*
49. *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.*
50. *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.*
51. *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.*
52. *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.*
53. *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.*
54. *Bharati A, RA KS. Crime prediction and analysis using machine learning. Int Res J Eng Technol (IRJET). 2018;5:1037–42.*
55. *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.*
56. *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.*
57. *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.*
58. *Sharma H, Kumar S. A survey on decision tree algorithms of classification in data mining. Int J Sci Res. 2016;5(4):2094–7.*
59. *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.*
60. *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.*
61. *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.*
62. *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.

63. *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.*
64. *Yerpude P, Gudur V. Predictive modelling of crime dataset using data mining. Int J Data Min Knowl Manag Process. 2020;7:83–99.*
65. *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.*