International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparative Study of Ml-Based Crime Prediction System

Riyaz Jamadar[1], Kunal Deshmukh[2], Niranjan Deshmukh[3], Aniket Kadale[4], Aditya Padwal[5].

[1]Assitant Professor, Department Of Information Technolgy, Aissms's Institute Of Information Technology, Pune-411001, India.
[2] BE (Information Technology), Aissms's Institute Of Information Technology, Pune-411001, India.

ABSTRACT:

In this era of recent times, crime has become an evident way of making people and society in trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyse and response ahead this type of criminal activity, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with the time period and location where it has happened.So far, this study reveals a global surge in crime rates, sparking concerted efforts to devise effective prediction tools. Synthesizing insights from diverse research papers on crime prediction, the first introduces a novel Naïve Bayes approach with 93.07% accuracy. Building on this, the second refines the K-Nearest Neighbor algorithm, crucial for proactive law enforcement in India. The third advances with convolutional neural networks, outperforming traditional methods.Practical applications include a Mumbai safety app (fourth), an optimized K-means algorithm (fifth), and a fairness-focused GANs architecture in Bogotá (sixth). The seventh introduces an integrated LSTM-ST-GCN model for predicting theft crimes, showcasing advanced predictive capabilities.

Lastly, the eighth conducts a systematic literature review, emphasizing the imperative for further research in spatio-temporal crime prediction.

In conclusion, this review provides a unified perspective on crime prediction advancements, traversing diverse methodologies and delineating strengths, acknowledging limitations, and suggesting avenues for future research.

Keywords: Crime prediction, Machine learning algorithms, Naïve Bayes K-Nearest Neighbor (KNN), Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Random Forest, Decision Trees, Spatio-temporal analysis, Predictive modeling**,** Law enforcement**,** Public Safety**,**Data Analysis**,** Comparative study**,** Literature review.

INTRODUCTION:

Crime prediction has become a pressing concern in contemporary society due to its significant impact on public safety and societal well-being. With the proliferation of criminal activities worldwide, there is a growing need for effective tools and methodologies to anticipate and prevent criminal incidents. In response to this challenge, researchers and law enforcement agencies have increasingly turned to machine learning (ML) algorithms to analyze crime data and forecast future criminal behavior.

The "Comparative Study of ML-Based Crime Prediction System" explores various ML techniques for crime prediction, aiming to enhance our understanding of crime patterns and improve proactive law enforcement strategies. By leveraging large datasets obtained from sources like Kaggle, researchers have endeavored to develop predictive models capable of identifying high-risk areas, predicting the occurrence of specific crimes, and ultimately assisting law enforcement agencies in allocating resources more efficiently.

This study synthesizes insights from a comprehensive literature review and evaluates the performance of different ML algorithms, including Naïve Bayes, K-Nearest Neighbor (KNN), Convolutional Neural Networks (CNN), and others. Through a systematic analysis of the strengths and limitations of each approach, the research aims to provide valuable guidance for selecting the most suitable ML algorithms for crime prediction tasks in diverse contexts.

Furthermore, the study identifies key challenges and opportunities in the field of crime prediction, emphasizing the importance of addressing issues such as data quality, algorithm robustness, and fairness in predictive modeling. By shedding light on the current state of ML-based crime prediction systems and highlighting areas for future research, this comparative study contributes to the ongoing efforts to enhance public safety and combat criminal activities worldwide.

LIMITATIONS:

1. Naïve Bayes Approach:

Assumes independence between features, which may not always hold in real-world scenarios. This assumption can lead to oversimplification, affecting the accuracy of predictions.

2. K-Nearest Neighbor Algorithm:

Sensitivity to outliers and noise in the data can result in less robust predictions. The choice of an optimal value for 'k' (number of neighbors) can also impact prediction accuracy.

3. Convolutional Neural Networks (CNN):

Requires a large amount of labeled data for effective training, and the complex architecture may lead to computationally intensive processes. Interpretability can be a challenge due to the "black-box" nature of deep learning models.

4. Optimized K-Means Algorithm:

Dependence on the initial choice of centroids can lead to suboptimal solutions. The algorithm is also sensitive to the number of clusters specified, which may not always be known beforehand.

5. Conditional Generative Adversarial Networks (GANs) Architecture:

Training GANs can be challenging and unstable. Achieving the right balance between accuracy and fairness may require careful fine-tuning, and the model's performance may be influenced by the quality and representativeness of the training data.

6. Integrated LSTM-ST-GCN Model:

Training complex models like LSTM-ST-GCN requires significant computational resources. Additionally, capturing all relevant spatial-temporal patterns in crime data may be challenging, and the model's performance may vary across different urban settings.

7. Spatio-Temporal Crime Hotspot Detection and Prediction (Literature Review):

The limitations of the reviewed models are context-dependent, and generalizing findings across diverse datasets and geographical regions may be challenging. The effectiveness of spatio-temporal models can be influenced by the granularity and quality of available data.

So from the above study, we understand that there are various limitations to ML algorithms. Hence, we can use **Support Vector Machines (SVM), Random Forest, and Decision Trees** are robust machine learning algorithms with distinct advantages in the realm of crime prediction. SVMs stand out for their prowess in handling non-linear decision boundaries, crucial for capturing intricate relationships within crime data. Their effectiveness in high-dimensional spaces makes them well-suited for scenarios involving multiple crime features. Additionally, SVMs prioritize margin optimization, enhancing generalization to unseen data. Random Forest, as an ensemble method, excels in mitigating overfitting by combining predictions from multiple decision trees. This approach not only enhances model robustness but also provides valuable insights into feature importance, aiding the identification of significant predictors for crime occurrences. Furthermore, Random Forest's reduced sensitivity to outliers contributes to stable predictions in the presence of noisy data. Decision Trees, known for their interpretability, offer transparency in understanding prediction logic, a crucial aspect for gaining insights into the factors influencing crime prediction. Their natural ability to model non-linear relationships and flexibility in handling diverse data distributions make Decision Trees a valuable choice in crime prediction tasks. The selection among these algorithms depends on the specific characteristics of the crime dataset and the objectives of the prediction task.

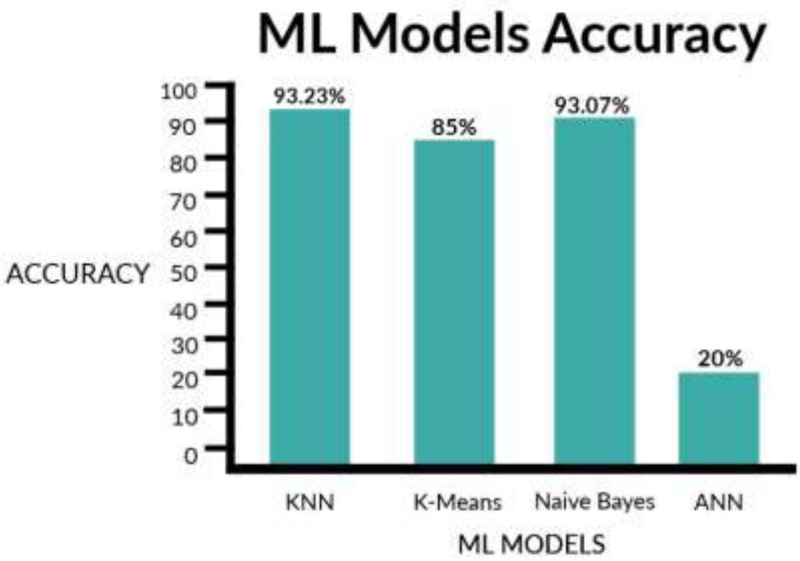
LITERATURE SURVEY:

Author	Paper Name	Abstract
1) Kanimozhi N 1, Keerthana N V2, Pavithra G S3, Ranjitha G4, Yuvaranis	Crime Type and Occurrence Prediction Using Machine Learning Algorithms	This paper introduces a machine learning approach for crime prediction, using existing data and considering temporal and spatial factors. The proposed Naïve Bayes classification system exhibits a significant accuracy improvement (93.07%) compared to current models. The study recommends exploring alternative classification models and examining the correlation between neighborhood income levels and crime rates, offering insights for effective crime-combatting strategies by law enforcement and policymakers.
2) Akash Kumar, Aniket Verma, Gandhali Shinde.	Crime Prediction Using K-Nearest Neighboring Algorithm	This paper proposes a K-Nearest Neighbor (KNN) algorithm for crime prediction in India. Utilizing historical data and geographical parameters, the study provides law enforcement with a proactive tool for identifying crime hotspots. After preprocessing and feature importance analysis, the optimized

			KNN algorithm shows a significant improvement in accuracy. The adaptable nature of KNN to evolving crime patterns underscores its value, offering valuable insights for more robust law enforcement strategies. The research highlights KNN's effectiveness in crime prediction, suggesting future exploration for expanding attributes and developing advanced classification algorithms to enhance accuracy and public safety.
3)	Atharva Deshmukh, Sourab Banka, Sean Brunzo, Sana Shaikh, Amiya Kumar Tripathy	Safety App: Crime Prediction Using GIS	This research tackles rising crime rates in India, exacerbated by technology and social media impact. It explores systematic methods for classifying crime patterns, emphasizing the effectiveness of the K-means algorithm. The study focuses on predicting high-crime regions and identifying crime-prone age groups. Notably, it introduces an optimized K-means algorithm to enhance efficiency. By addressing these aspects, the research aims to significantly advance crime analysis and prediction methodologies in India. Keywords: crime, clustering, optimized K-means algorithm.
4)	Krishnendu S.G, Lakshmi P.P, Nitha L	Crime Analysis and Prediction using Optimized K-Means Algorithm	This research addresses rising crime rates in India amidst technological and social media influences. It explores systematic methods for classifying crime patterns, emphasizing limitations in existing clustering algorithms. The study highlights the efficacy of the K-means algorithm for predicting high-crime regions and discerning crime-prone age groups. Introducing an optimized K-means algorithm for efficiency, the research aims to advance crime analysis and prediction methodologies in India. Keywords: crime, clustering, optimized K-means algorithm.
5)	Fatih Ilhan1,3 , Selim F. Tekin1,3 ve Bilgin Aksoy2,3 1Elektrik ve Elektronik Mühendisligi Bölümü, ˘ ˙Ihsan Dogramacı Bilkent Üniversitesi	Spatio-Temporal Crime Prediction with Temporally Hierarchical Convolutional Neural Networks	This paper introduces a novel deep learning model using convolutional neural networks for spatio-temporal crime prediction. Employing a temporally hierarchical structure and channel projection, the model captures temporal patterns and separate influences of crime events on future crime risk. Results on Chicago and Los Angeles crime datasets show significant performance improvement over classical methods.
6)	Shuyu Yao, Ming Wei, Lingyu Yan1*, Chunzhi Wang, Xinhua Dong, Fangrui Liu, Ying Xiong	Prediction of Crime Hotspots based on Spatial	This paper addresses the critical aspect of crime prediction in contemporary policing, recognizing the limitations inherent in relying solely on historical crime data. Introducing the Random Forest algorithm as a machine learning solution to enhance crime hotspot prediction, the research extensively reviews previous studies in crime prediction, emphasizing a place-centered paradigm and incorporating diverse data types, such as Points of Interest (POI) and demographic data. Utilizing crime data from San Francisco, the study categorizes areas into distinct hotspots and non-hotspots, employing the Random Forest model with covariates for prediction. Analyzing historical crime trends and presenting experimental results comparing various models, including Naive Bayes and logistic regression, the paper underscores the superiority of the Random Forest model with covariates in predicting crime hotspots. It also emphasizes the stability of crime hotspots in spatial terms and advocates for future research to incorporate temporal factors.

7)	Pratibha,Akanksha Gahalot, Uprant	Crime Prediction and Analysis	This paper addresses the challenge of rising crime rates, emphasizing the need for effective prediction tools. It explores diverse machine learning algorithms, including Extra Tree Classifier, K-Nearest Neighbor, SVM, Decision Tree Classifier, and ANN. The research covers an introduction, related work review, detailed methodology discussions, practical implementation insights, comprehensive result discussions, and a conclusion summarizing key findings. The paper contributes to the discourse on crime prediction and analysis, suggesting potential avenues for future research.
8)	Christian Urcuqui, Juan Moreno, Carlos Montenegro, Álvaro Riascos, Mateo Dulce	Accuracy and Fairness in a Conditional Generative Adversarial Model of Crime Prediction	This paper introduces a conditional GANs architecture using ConvLSTM neural networks for predicting violent robberies in Bogotá, Colombia, outperforming existing models. The authors address fairness concerns by conducting a calibration test based on protected variables, revealing bias linked to residential income. They propose a fairness-accuracy balancing technique to mitigate bias while maintaining accuracy.

DISCUSSION:



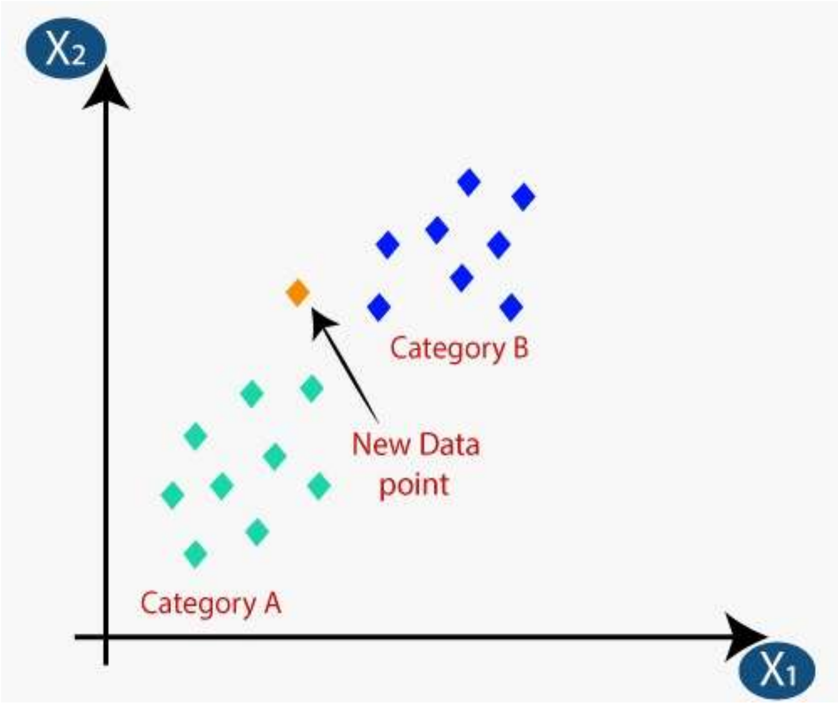
The bar graph displays the accuracy of several machine learning models, with the data sourced from a comprehensive review of various research papers. Among the models evaluated, K-Nearest Neighbors (KNN) and Naive Bayes exhibit the highest accuracy, with KNN achieving an impressive accuracy rate of 93.23%, closely followed by Naive Bayes at 93.07%. In contrast, K-Means demonstrates a respectable accuracy of 85%, showcasing its utility in certain applications. However, the Artificial Neural Network (ANN) lags significantly behind, with an accuracy of just 20%. These findings underscore the importance of selecting the right machine learning model for specific tasks, as performance can vary significantly across different algorithms.

It's important to note that while these models showcase varying degrees of accuracy, they may also come with inherent drawbacks. To further explore their strengths and weaknesses, our comparative study will incorporate Support Vector Machines (SVM), Random Forest, and Decision Trees. This comparative analysis will provide a more comprehensive understanding of the performance and suitability of these machine learning models for different tasks.

1. K-Nearest Neighbors (KNN):

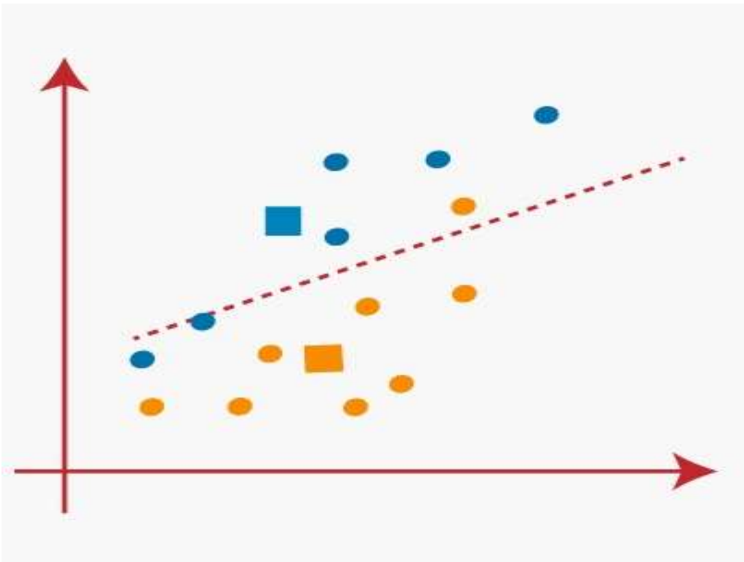
KNN is a simple and intuitive algorithm used for classification and regression tasks.It operates based on the principle that similar data points are close to each other in the feature space.In classification, when a new data point needs to be classified, KNN identifies the K nearest data points (neighbors) based on some distance metric (such as Euclidean distance) and assigns the majority class among them to the new data point.

KNN doesn't involve a training phase; instead, it memorizes the entire training dataset. Therefore, its prediction phase can be computationally expensive for large datasets.One of the key hyperparameters in KNN is the value of K, which determines the number of neighbors considered for classification.



2. K-means Clustering:

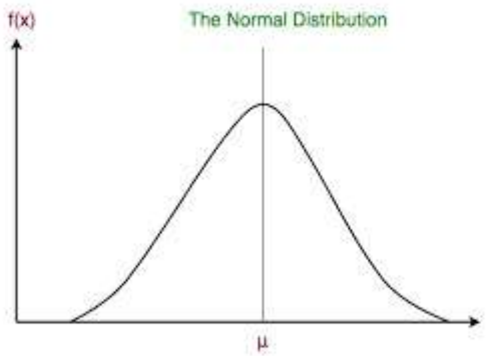
K-means is an unsupervised machine-learning algorithm used for clustering tasks.It aims to partition a dataset into K clusters, where each data point belongs to the cluster with the nearest mean (centroid).The algorithm iteratively assigns data points to the nearest centroid and then updates the centroids based on the mean of the points assigned to each cluster. K-means clustering is sensitive to the initial choice of centroids, and different initializations can lead to different final cluster assignments.The algorithm converges when the centroids no longer change significantly between iterations or when a maximum number of iterations is reached.



3. Naïve Bayes:

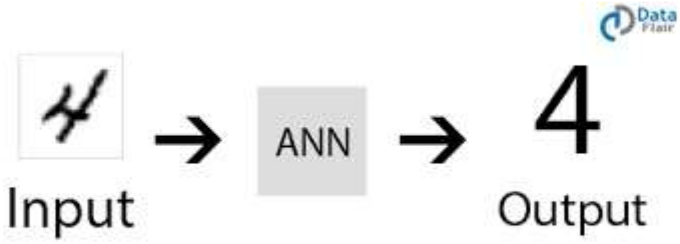
Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem with the "naive" assumption of feature independence.Despite its simplistic assumption, Naïve Bayes often performs well in practice, especially for text classification tasks.It calculates the probability of each class given a set of input features and selects the class with the highest probability as the predicted class.

Naïve Bayes models are efficient and require a small amount of training data to estimate the parameters.However, the assumption of feature independence may not hold true in many real-world scenarios, which can affect the accuracy of the model.



4. Artificial Neural Networks (ANN):

ANN is a computational model inspired by the structure and function of the human brain's neural networks. It consists of interconnected nodes (neurons) organized in layers: an input layer, one or more hidden layers, and an output layer.Each connection between neurons is associated with a weight that is adjusted during the training process to minimize the difference between predicted and actual outputs. ANNs are capable of learning complex patterns and relationships in data, making them suitable for a wide range of tasks, including classification, regression, and pattern recognition.



CONCLUSION:

In conclusion, our study highlights the importance of utilizing machine learning (ML) algorithms for crime prediction. While various ML methods have been explored, including Naïve Bayes, K-Nearest Neighbor (KNN), and Convolutional Neural Networks (CNN), each has its limitations. However, Support Vector Machines (SVM), Random Forest, and Decision Trees emerge as robust options for crime prediction due to their distinct advantages.

SVMs excel in handling complex relationships within crime data, while Random Forest mitigates overfitting and provides insights into feature importance. Decision Trees offer transparency in understanding prediction logic, crucial for gaining insights into crime factors. The choice among these algorithms depends on specific dataset characteristics and prediction objectives.

Our comparative analysis, along with insights from existing research papers, underscores the need for further exploration of ML techniques in crime prediction. By addressing the limitations and leveraging the strengths of different algorithms, we can enhance the effectiveness of predictive models, ultimately contributing to improved public safety strategies.

Overall, our study advocates for continued research in ML-based crime prediction, emphasizing the importance of selecting suitable algorithms tailored to the unique characteristics of crime datasets and prediction tasks.

REFERENCES:

[1] H. Su, K. Zheng, J. Huang, H. Jeung, L. Chen, and X. Zhou, “Crowdplanner : a crowd-based route recommendation system,” in Proceedings of the 30th IEEE International Conference on Data Engineering (ICDE’14), pp.1144–1155, IEEE, Chicago, USA, April 2014.

[2] V. Arnaboldi, M. Conti, and F. Delmastro, “Implementation of CAMEO: a context-aware middleware for opportunistic mobile social networks”, in Proceedings of the IEEE International Sym posium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM’11), pp.1–3, Lucca, Italy, June 2011.

[3] Firoz Ahmad, Md Meraj Uddin, Laxmi Goparaju, ” Role of Geospatial Technology in Crime Mapping: A Perspective View of India”, in World Scientific News Journal 88(2), Apirl 2017, pp.211-226.

[4] M. Nagarajan, K. Gomadam, A. P. Sheth, A. Ranabahu, R. Mutharaju, and A. Jadhav, “Spatiotemporal-thematic analysis of cit izen sensor data: challenges and experiences,” in Web Information Systems Engineering— WISE2009, vol.5802 of Lecture Notes in Computer Science, pp. 539–553, Springer, Berlin, Germany, 2009.

[5] S.C.M. de S Sirisuriya, "A Comparative Study on Web Scraping" in Proceedings of 8th International Research Conference, KDU, Nov 2015, pp.135-140.

[6] Felix Mata, Miguel Torres-Ruiz, Giovanni Guzm ´ an, Rolando Quin- ´tero, Roberto Zagal-Flores, Marco Moreno-Ibarra, and Eduardo Loza, "A Mobile Information System Based on Crowd-Sensed and Official Crime Data for Finding Safe Routes: A Case Study of Mexico City", Hindawi Publishing Corporation Mobile Information System, Article ID 8068209, pp.3-5, 2016.

[7] Varvara Ingilevicha, Sergey Ivanov, "Crime rate prediction in the urban environment using social factors", 7th International Young Scientists Conference on Computational Science, YSC2018, 02-06 July2018, Heraklion, Greece.

[8] Hitesh Kumar Reddy, Bhavna Sainia, Ginika Mahajana. "Crime Prediction and Monitoring Framework Based on Spatial Analysis". International Conference on Computational Intelligence and Data Science (ICCIDS), Volume 132, 2018, Pages 696-705.

[9] Alexandros Belesiotis, George Papadakis, Dimitrios Skoutas . "An alyzing and Predicting Spatial Crime Distribution Using Crowd sourced and Open Data". ACM Transactions on Spatial Algorithms and Systems, Vol. 3, No. 4, Article 12 (2018).

[10] Isha Pradhan, Katerina Potika, Magdalini Eirinaki and Petros Potikas. "Exploratory Data Analysis and Crime Prediction for Smart Cities", Proceedings of the 23rd International Database Applications and Engineering Symposium, June 2019, Article No: 4, Pages 1–9.

[11] Aman Kumar, Nikhil Tiwari, Prakhar Gupta, S.N.Rajan. "WOMEN CRIME PREDICTION", International Research Journal of Engi neering and Technology (IRJET), Volume: 04, Apr -2017.

[12] Ginger Saltos, Mihaela Cocea. "An Exploration of Crime Prediction Using Data Mining on Open Data", Article in International Journal of Information Technology and Decision Making, May 2017.

[13] Alkesh Bharati, Sarvanguru RA.K, "Crime Prediction and analysis using Machine Learning", International Research Journal of Engi neering and Technology (IRJET), Volume: 05, Sep 2018.

[14] Chao Huang, Junbo Zhang, Yu Zheng, Nitesh Chawla, "Deepcrime: Attentive Hierarchical Recurrent Networks for Crime Prediction", Proceedings of the 27th ACM International Conference on Informa tion and Knowledge Management, October 2018, pp.1423–1432.

[15] http://mahacid.com/monthly-clime-review.html last accessed on 31-12-2019.

[16] http://www.mhpolice.maharashtra.gov.in/Citizen/MH/ last accessed on 31-12-2019.

[17] http://ncrb.gov.in/ last accessed on 31-12-2019.

[18] https://data.gov.in/ last accessed on 31-12-2019.

[19] https://www.diva-gis.org/ last accessed on 31-12-2019.

[20] https://timesofindia.indiatimes.com/city/mumbai/mumbai two-arrested-for-stealing-goods-from-moving-train coaches/articleshow/71788161.cms last accessed on 31-12-2019.

[21] Kim, Suhong, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri. "Crime Analysis Through Machine Learning." In 2018 IEEE 9th Annual

Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 415-420. IEEE, 2018.

[22] Shah, Riya Rahul. "Crime Prediction Using Machine Learning." (2003).

[23] Lin, Ying-Lung, Tenge-Yang Chen, and Liang-Chih Yu. "Using machine learning to assist crime prevention." In 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 1029-1030. IEEE, 2017.

[24] M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 2016.

[25]Crime Prediction Using Machine Learning Sacramento Stateathena.ecs.csus.edu › ~shahr › progress_report by RR Shah - 2003.

[26] Williams, Matthew L., Pete Burnap, and Luke Sloan. "Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns." The British Journal of Criminology 57, no. 2 (2017): 320-340.

[27] Agarwal, Shubham, Lavish Yadav, and Manish K. Thakur. "Crime Prediction Based on Statistical Models." In 2018 Eleventh International Conference on Contemporary Computing (IC3), pp. 1-3. IEEE, 2018.

[28] Nakaya, Tomoki, and Keiji Yano. "Visualising crime clusters in a spacetime cube: An exploratory dataanalysis approach using space time kernel density estimation and scan statistics." Transactions in GIS 14, no. 3 (2010): 223-239

[29]Khushabu A. Bokde,Tiksha P. Kakade, Dnyaneshwari S. Tumsare, ChetanG.Wadhai: Crime Analysis Using K Means Clustering(2018).https://www.ijert.org/crime analysis-using-k-means-clustering.

[30] M. DeLisi, A. Kosloski, M. Sween, E. Hachmeister, M. Moore, and A. Drury, "Murder by numbers: Monetary costs imposed by a sample of homicide offenders," The Journal of Forensic Psychiatry Psychology, pp. 501–513, 2010.

[31] P. Buonanno, “Broken windows,” Atlantic monthly, no. 3, pp. 29–38, 1982.

[32] R. Wright and S. Decker, “Burglars on the job: Streetlife and residential break-ins,” UPNE, no. 3, pp. 29–38, 1996.

[33] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, “Self-exciting point process modeling of crime,” Journal of the American Statistical Association, vol. 106, no. 493, pp. 100–108, 2011.

[34] G. Mohler, “Marked point process hotspot maps for homicide and gun crime prediction in chicago,” International Journal of Forecasting, vol. 30, no. 3, pp. 491–497, 2014.

[35] S. Xingjian, Z. Chen, D. Yeung, W. Wong, and W. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” In Advances in neural information processing systems, pp. 802–810, 2015.

[36] R. S. Dulce, M and R. A, “Efficient allocation of law enforcement resources using predictive police patrolling,” In Advances in neural information processing systems, 2018.

[37] S. Barocas and A. D. Selbst, “Big data’s disparate impact,” California Law Review, 2016.

[38] A. Stec and D. Klabjan, “Forecasting crime with deep learning,” arXiv preprint arXiv:1806.01486, 2018.

[39] D. Saxena and J. Cao, “D-gan: Deep generative adversarial nets for spatio-temporal prediction,” arXiv preprint arXiv:1907.08556, 2019.

[40] K. M. Jonietz D, “Towards modeling geographical processes with generative adversarial networks (gans),” In 14th International Conference on Spatial Information Theory (COSIT 2019), 2019.

[41] M. Hardt, E. Price, and N. Srebro, “Equality of opportunity in super vised learning,” In Advances in neural information processing systems, pp. 3315–3323, 2016.

[42] K. Lum and W. Isaac, “To predict and serve?,” Significance, pp. 14–19, 2016.

[43] G. Mohler, R. Raje, J. Carter, B. Valasik, M, and J, “A penalized likelihood method for balancing accuracy and fairness in predictive policing,” In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC).IEEE, 2454-2459.

[44] R. R. W. F. K. J. Pleiss, G and K. Q. Weinberger, “On fairness and calibration,” 31st Conference on Neural Information Processing Systems

(NIPS 2017), Long Beach, CA, USA., https://arxiv.org/abs/1709.02012.

[45] A. Kumar, S. Sarawagi, and U. Jain, “Trainable calibration measures for neural networks from kernel mean embeddings,” In International Conference on Machine Learning, 2805-2814.

[46] P. N. de Colombia, “Registros administrativos de delitos y denuncias. sistema de información estadística, delincuencial, contravencional y operativa (siedco),” 2020. data retrieved from SIEDCO, https://scj.gov. co/en/oficina-oaiee/bi/seguridad_convivencia/siedco.

[47] C. D. C. De Bogota et al., “Perfil económico y empresarial: Localidad chapinero,” 2006.

[48] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, “Generative adversarial nets,” In Advances in neural information processing systems, 2672-2680.

[49] D. Ensign, S. A. Friedler, S. Neville, C. Scheidegger, and S. Venkata subramanian, “Runaway feedback loops in predictive policing,” arXiv preprint arXiv:1706.09847, 2017

[50] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pent land, “Once upon a crime: towards crime prediction from demographics and mobile data”,IEEE,Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 427-434.

[51] Ubon Thansatapornwatana,”A Survey of Data Mining Techniques for Analyzing Crime Patterns”, Second Asian Conference on Defense Technology ACDT, IEEE,Jan 2016, pp. 123–128.

[52] H. Adel, M. Salheen, and R. Mahmoud, ”Crime in relation to urban design. Case study: the greater Cairo region,” Ain Shams Eng. J., vol. 7, no. 3, pp. 925-938, 2016.

[53] J. L. LeBeau, ”The Methods and Measure of Centrography and the spatial Dynamics of Rape” Journal of Quantitative Criminology, Vol.3, No.2, pp.125-141, 1987.

[54] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, Alex Pentland.”Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data”,in ACM International Conference on Multimodal Interaction (ICMI 2014).

[55] Shiju Sathyadevan,Devan M. S.,Surya S Gangadharan, First,”Crime Analysis and Prediction Using Data Mining” International Conference on Networks Soft Computing (ICNSC), 2014.

[56] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,”Crime pattern detection,analysis and predic tion,International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2017.

[57] Amanpreet Singh,Narina Thakur,Aakanksha Sharma,”A review of su pervised machine learning algorithms”,3rd International Conference on Computing for Sustainable Global Development,2016.

[58] Bin Li,Yajuan Guo,Yi Wu,Jinming Chen,Yubo Yuan,Xiaoyi Zhang,”An unsupervised learning algorithm for the classification of the protection device in the fault diagnosis system”,in China International Conference on Electricity Distribution (CICED),2014.

[59] Varshitha D N Vidyashree K P,Aishwarya P Janya T S,K R Dhananjay Gupta Sahana R,”Paper on Different Approaches for Crime Prediction system”,International Journal of Engineering Research Technology (IJERT), ISSN: 2278-0181, 2017

[60] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, ”An experimental study of classification algorithms for crime prediction,” Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.

[61] Malathi. A,Dr. S. Santhosh Baboo,”An Enhanced Algorithm to Predict a Future Crime using Data Mining”,International Journal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011.

[62] T. Beshah and S. Hill, ”Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia,” Proc. of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.

[63] K.B.S. Al-Janabi, “A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-Means Mining Algorithm,” in Journal of Kufa for Mathematics and Computer, Vol. 1, No. 3, 2011, pp. 8-24.

[64] A. Malathi, S.S. Baboo, “An Enhanced Algorithm to Predict a Future Crime using Data Mining,” in International Journal of Computer Applications, Vol. 21, 2011, pp. 1-6.

[65] C.H. Yu, “Crime Forecasting Using Data Mining Techniques,” in 11th International Conference on Data Mining Workshop, 2011, pp. 779-786.

[66] Rodriguez-Galiano V F, Ghimire B, Rogan J, et al. An assessment of the effectiveness of a random forest classifier for land-cover classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012,67:93-104.

[67] Ma H J, Gao X H, Gu X T. Land use/land cover classification of complex terrain areas supported by random forest method[J]. Journal of Geo-information Science, 2019,21(3):359-371.

[68] Lu R, Li LY. Crime prediction model based on random forest[J]. Journal of China Criminal Police College, 2019(3):108-112.

[69] BogomolovA, Lepri B, Staiano J. Once upon a crime: Towards crime prediction from demographics and mobile data[C]. International Conference on Multimodal InteractionACM, 2014:427-434

[70] Levine N. Commentary the "Hottest" part of a hotspot: Comments on "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime"[J]. Security Journal, 2008,21:295-302.

[71] Eck, J., Chainey, S., Cameron, J. and Wilson, R., 2005. Mapping crime: Understanding hotspots.

[72] Qin X, Liu L. Evaluating the relationships of bus transit with street and off- street robberies[J]. Professional Geographer, 2015,68(2):1-11.

[73] Hart T C, Miethe T D. Street robbery and public bus stops: A case study of activity nodes and situational risk [J]. Security, 2014,27(2):180-193.

[74] Blair L, Wilcox P Eck J. Facilities, opportunity, and crime: An exploratory analysis of places in two urban neighborhoods[J]. Crime Prevention & Community Safety, 2017,19(1):1-21.

[75] Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P. and Tita, G.E., 2011. Self-exciting point process modeling of crime. Journal

of the American Statistical Association, 106(493), pp.100-108.

[76] Wang, H., Kifer, D., Graif, C. and Li, Z., 2016, August. Crime rate inference with big data. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 635-644). ACM.

[77] Hoeben, E. M., Bernasco, W., Weerman, F. M., Pauwels, L., & van Halem, S. (2014). The spaceetime budget method in criminological research. Crime Science, 3(1),12.

[78] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1):5-32.[1] S. Hussain et al., “Impact of crime and corruption on gdp per capita an

empirical analysis of cross-country data.” Pakistan Journal of Criminol ogy, vol. 10, no. 2, 2018.

[79] J. Wang, J. Hu, S. Shen, J. Zhuang, and S. Ni, “Crime risk analysis through big data algorithm with urban metrics,” Physica A: Statistical Mechanics and its Applications, vol. 545, p. 123627, 2020.

[80] J. Chin and C. Bürge, “Twelve days in xinjiang: how china’s surveillance state overwhelms daily life,” Wall Street Journal, vol. 19, 2017