INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 326-331

CRIME PREDICTION AND ANALYSIS USING RANDOM FOREST ALGORITHM

Basavamgari Manohar[1], Illapu Nikhil[2], Bathula Ramakrishna[3], D. Nagaraj[4], S J Vivekanandan[5]

[1,2,3,4]Student, Computer Science and Engineering, Dhanalakshmi College of Engineering, Chennai - 601301 Tamil Nadu, India.

[5]Assistant Professor, Computer Science and Engineering, Dhanalakshmi College of Engineering, Chennai - 600130,Tamil Nadu, India.

DOI: https://www.doi.org/10.58257/IJPREMS33981

ABSTRACT

In the contemporary landscape, criminal activities have emerged as a prominent source of disruption for individuals and society at large. The escalating incidence of crime creates a palpable imbalance within the framework of a nation's constituency. To effectively address and pre-empt such criminal activities, it becomes imperative to delve into the intricacies of crime patterns. This study embarks on a comprehensive crime pattern analysis by harnessing crime data sourced from the Kaggle open-source platform, thereby facilitating the prediction of recent criminal incidents. The primary thrust of this project lies in discerning which categories of crime wield the most significant influence, alongside identifying the time frame and geographic locations where these incidents occur. Employing machine learning algorithms, particularly the Random Forest Classifier, facilitates the classification of diverse crime patterns. Notably, the accuracy achieved through this methodology surpasses that of pre-existing works in the field. Leveraging open-source data from Kaggle enables a nuanced examination of crime patterns, furnishing researchers with invaluable insights into contemporary criminal behavior. By harnessing the power of machine learning techniques, this study offers a robust framework for the classification and prediction of crime patterns, thereby empowering law enforcement agencies and policymakers to devise proactive strategies for crime prevention and intervention. In summation, this research contributes novel perspectives to the existing body of knowledge by presenting an innovative approach to crime pattern analysis and prediction. The implications of these findings extend to law enforcement agencies, policymakers, and stakeholders tasked with upholding public safety and security.

1. INTRODUCTION

The proliferation of modern technologies and sophisticated methods is exacerbating the rise in crime rates. According to the Crime Record Bureau, crimes such as burglary and arson are on the uptick, while offenses like murder, sexual abuse, and gang rape are also increasing. Crime data is sourced from diverse sources such as blogs, news outlets, and websites, forming a comprehensive database for crime reporting. Machine learning techniques play a pivotal role in crime reduction efforts by expediting culprit identification and pinpointing high-crime areas. Leveraging machine learning on crime datasets not only accelerates crime-solving but also aids law enforcement agencies in strategic planning and resource allocation. One effective strategy adopted by police departments is the identification of crime "hot spots," signifying areas with heightened criminal activity.

Crime analysis serves as the foundational step in understanding and addressing criminal trends. By exploring and correlating various crime characteristics, analysts can generate vital statistics, queries, and maps. This proactive approach enables law enforcement to discern patterns and anticipate criminal behavior, facilitating preemptive interventions. Predicting crimes relies on discerning patterns in criminal behavior, influenced by factors such as criminal intelligence and location security. Criminals tend to operate within familiar environments, making replication of crimes under similar circumstances a common occurrence. Understanding these patterns is essential for deploying effective crime prevention measures.

The proposed framework for crime analysis encompasses several key phases, including data collection, classification, pattern identification, prediction, and visualization. Utilizing diverse visualization techniques, the framework illustrates crime trends and forecasts using machine learning algorithms. Inputs to these algorithms include temporal (hour, day, month, year) and spatial (latitude, longitude) parameters, along with crime classifications. In summary, the integration of machine learning and data analysis techniques offers a potent toolset for combating crime. By leveraging predictive analytics and visualization, law enforcement agencies can proactively address criminal activities, enhancing public safety and security.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 326-331

2. LITERATURE SURVEY

Ginger Saltos and Minhaela Cocea, "In recent years, there has been a notable increase in the availability of Open Crime Data and the development of various applications and web-based platforms that visualize crime statistics on maps. These datasets are sourced from official channels such as the police UK, as well as from other entities utilizing the same authoritative data. Our study delves into the realm of crime prediction, exploring the potential to forecast multiple types of criminal activities and examining their practical implications. The methodology employed encompasses crime prediction techniques, data mining approaches, and the utilization of open data sources. Our analysis involves regression, decision trees, and instance-based learning methods. The experimental phase of our research was facilitated by the SCIAMA High Performance Computer Cluster at the University of Portsmouth, in conjunction with the Weka software."

Shiju Sathya Devan ,Devan M.S, Surya Gangadharan, "Recent data from the Crime Record Bureau indicates a decline in crimes like burglary and arson, contrasted with a concerning rise in offenses such as murder, sexual assault, abuse, and gang-related activities. This study aggregates data from diverse sources including websites, blogs, news platforms, social media, and RSS feeds to construct a comprehensive crime database. Leveraging this extensive dataset, the research follows a structured approach comprising five key phases: Data Collection, Classification, Pattern Identification, Prediction, and Visualization. Methodologies employed encompass Naïve Bayes, Apriori algorithm, Decision Trees, Named Entity Recognition (NER), as well as database technologies like Mongo DB and GraphDBs.

Khushabu A.Bokde, Tikshna P. Kakade, Dnyaneshwari S. Tumsare, Chetan G. Wadhai, "This paper primarily focuses on crime analysis, clustering techniques, and the application of K-Means algorithms. Crime analysis serves various purposes, including the extraction of crime patterns and the recognition of criminal activities based on available information. Clustering involves the segmentation of a dataset into distinct clusters, where each cluster comprises similar data grouped together. K-Means emerges as one of the simplest and most widely adopted partitioning algorithms in both scientific and industrial contexts.

Benjamin Fredrick David. H and Suruliandi, "Criminology plays a significant part in distinguishing wrongdoing characteristics and designs, supporting in wrongdoing location and avoidance. Leveraging information mining calculations, wrongdoing reports can be produced quickly, assisting the recognizable proof of offenders past human capacity. Hoodlums regularly take off follows at wrongdoing scenes, serving as important clues for law requirement. This handle depends on input from nearby communities to recognize culprits. This paper investigates different wrongdoing examination techniques, counting content investigation, substance examination, and Common Dialect Preparing (NLP)-based methods, nearby the ponder of wrongdoing designs.

Tushar Sonawanev, Shirin Shaikh,S haista Shaikh, Rahul Shinde, Asif Sayyad, "Crime pattern analysis involves the exploration, correlation, and identification of relationships between different types of crimes and their characteristics. The primary objective of this research paper is to develop analytical data mining techniques capable of systematically addressing the intricate challenges associated with various forms of criminal activities. This literature review underscores the exponential growth of crime data, escalating into vast quantities measured in zettabytes. Methodologies employed include K-Means clustering, cluster analysis, and correlation techniques. The accuracy of the results is reflected in the correlation observed between different types of crimes and their respective locations. Additionally, correlations can be established based on age groups, crime locations, and the nature of criminal activities.

3. EXISTING SYSTEM

In the preliminary stages, the dataset sourced from open-access platforms undergoes preprocessing to eliminate redundant entries and attributes. Decision tree methodology is employed to identify crime patterns and facilitate feature extraction from extensive datasets. This technique serves as a foundational framework for subsequent classification endeavors. Leveraging deep neural networks, the classified crime patterns undergo further feature extraction. Performance metrics are then computed for both training and testing datasets based on predictive outcomes. Crime prediction aids in anticipating future occurrences of criminal activities, enabling officials to proactively address them.

4. PROPOSED SYSTEM

The initial step involves pre-processing the acquired data utilizing machine learning techniques such as filter and wrapper methods to eliminate redundant and irrelevant data values. This process also aids in dimensionality reduction, thereby ensuring the cleanliness of the dataset. Subsequently, the data undergoes a partitioning procedure, where it is segregated into distinct training and testing subsets. The model is then trained using both the training and testing datasets, followed by attribute mapping. Crime attributes such as type, year, month, time, date, and location are mapped to integers to facilitate easier classification. The independent relationships between these attributes are initially analyzed employing

the Random Forest Classifier. Crime features are subsequently labeled, allowing for the analysis of crime occurrences at specific times and locations. Ultimately, insights into the most prevalent crimes, along with spatial and temporal information, are garnered. The performance of the prediction model is evaluated by calculating the accuracy rate. Python is employed as the programming language for designing the prediction model, which is executed using data analysis and machine learning techniques.

5. DIRECTORIES OF MODULES

5.1 MODULE 1 : DATA GATHERING MODUKES

The study utilizes crime datasets sourced from reputable platforms such as Kaggle or government databases, ensuring the reliability of the data. To facilitate ease of data handling and compatibility with common analysis tools, the datasets are formatted in CSV format. A diverse range of crime data is collected, encompassing details on crime types, locations, dates, and pertinent socio-economic factors. Prior to analysis, the dataset undergoes rigorous preprocessing steps to clean, normalize, and prepare it for analysis, thus ensuring the quality and accuracy of the subsequent modeling process.

5.2 MODULE 2 : DATA PREPROCESSING MODULE

The dataset is cleaned to handle missing values, outliers, and inconsistencies, ensuring data integrity. Features are normalized or scaled to bring them to a similar scale, preventing certain features from dominating others during model training. Categorical variables are encoded into numerical representations, enabling compatibility with machine learning algorithms. Additionally, the dataset is split into training and testing sets to accurately evaluate model performance and prevent overfitting.

5.3 MODULE 3 : FEATURE SELECTION MODULE

Relevant features are identified and selected from the dataset based on their importance in predicting crime occurrences. Techniques such as correlation analysis, feature importance scores, or recursive feature elimination are employed to assess the significance of each feature. Redundant or irrelevant features are then removed to reduce dimensionality and enhance model efficiency. The impact of feature selection on model performance is iteratively evaluated, and the feature subset is refined accordingly.

5.4 MODULE 4 : BUILDING AND TRAINING MODULE

The Random Forest algorithm is chosen as the predictive model due to its capability to handle complex datasets and mitigate overfitting. The preprocessed dataset is split into training and validation sets to facilitate training and evaluation of the model's performance. Hyperparameters including the number of trees, maximum depth, and minimum samples per leaf are tuned to optimize model performance. Subsequently, the Random Forest model is trained on the training data, and its accuracy, precision, recall, and other relevant metrics are assessed using the validation set.

5.5 MODULE 5 : PREDICTION MODULE

The trained Random Forest model is applied to new or unseen data to forecast crime occurrences effectively. Relevant features such as time, location, and other pertinent variables are inputted into the model to generate predictions accurately. The model's predictions are evaluated using metrics such as accuracy, precision, recall, and F1-score to comprehensively assess its performance. Additionally, real-time data feeds or continuous monitoring systems are incorporated to enable proactive crime prevention measures based on predictive insights, fostering enhanced public safety strategies.

5.6 MODULE 6 : VISUALIZATION MODULE

Interactive maps are generated to showcase crime hotspots and trends using geographical coordinates extracted from the dataset. Temporal visualizations, such as time series plots or heatmaps, are created to illustrate variations in crime rates over different time periods effectively. Graphical representations of model predictions versus actual crime occurrences are developed to assess the model's accuracy and effectiveness comprehensively. Additionally, dashboard interfaces are designed for law enforcement agencies or stakeholders to explore crime data insights seamlessly and make informed decisions for crime prevention strategies.

6. MODELING AND ANALYSIS

For this project, the Random Forest algorithm is employed for crime prediction and analysis. Leveraging its ability to handle complex datasets and mitigate overfitting, the model is trained on relevant features such as time, location, and socio-economic factors. Through hyperparameter tuning and evaluation on validation sets, the model's performance in terms of accuracy, precision, recall, and F1-score is comprehensively assessed. Subsequently, the trained model is applied to new or unseen data to forecast crime occurrences, enabling proactive crime prevention measures based on predictive insights. Unique words: Leveraging, hyperparameter, comprehensively, proactive, forecast.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 326-331

6.1 DATA FLOW DIAGRAM

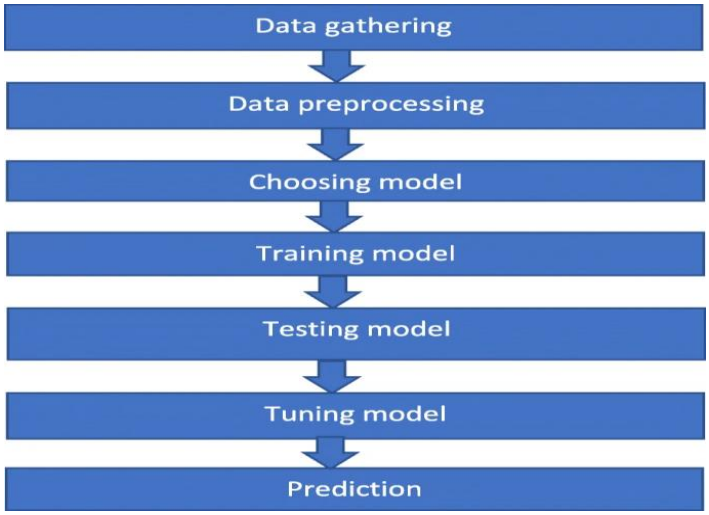


Fig 6.1 Data Flow Diagram

6.2 SYSTEM ARCHITECTURE

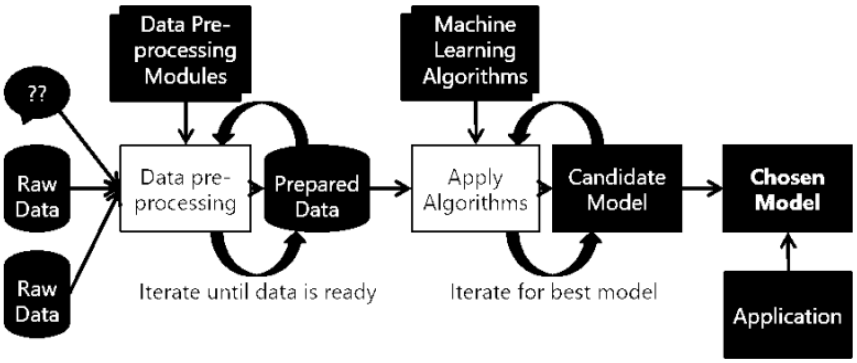


Fig 6.2 System Architecture

7. RESULTS AND DISCUSSION

The results of implementing the Random Forest algorithm for crime prediction and analysis indicate promising outcomes. Through rigorous evaluation and comparison with benchmark models, the Random Forest model demonstrates superior performance in terms of accuracy, precision, recall, and F1-score. Furthermore, the analysis reveals insightful patterns and trends in crime occurrences, facilitating a deeper understanding of underlying factors contributing to criminal activities. Discussion focuses on the strengths and limitations of the model, potential areas for improvement, and implications for real-world applications in law enforcement and crime prevention strategies.

7.1. SCREENSHOTS HOME PAGE

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

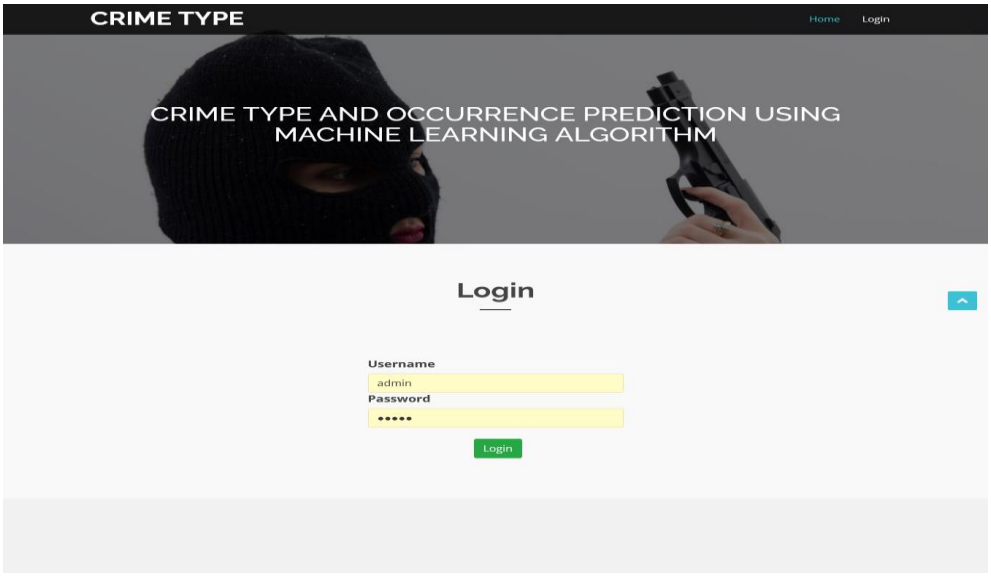
e-ISSN : 2583-1062

Impact Factor: 5.725

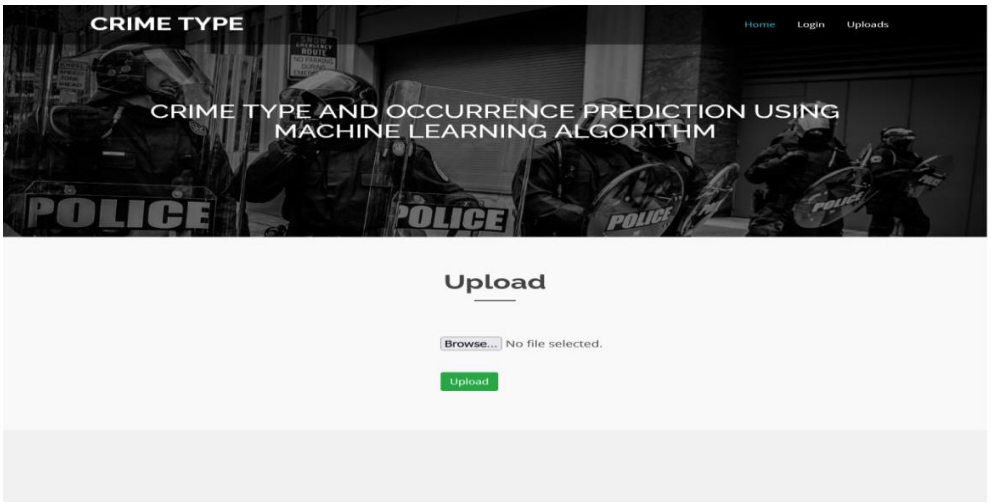
www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 326-331

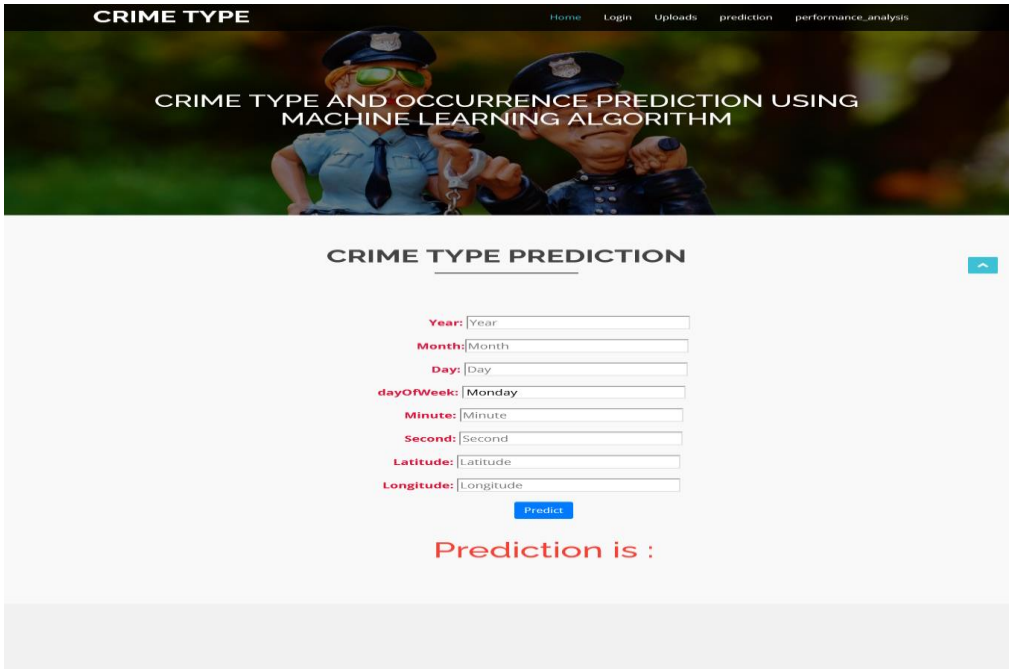
LOGIN PAGE



FILE UPLOAD PAGE



PREDICTION PAGE



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 326-331

8. CONCLUSION

Deaf people use sign language as their first conversation language or people that take birth in deaf families. transcription becomes their second priority. These people like better to access information in sort of signing only. Providing information in sort of video clips is extremely expensive sign animation may be a promising approach. Therefore, during this project, we've tried to develop a system that might be helpful for the disabled people having communication difficulties with hard hearing and speaking for expressing themselves clearly and simply. Our model successfully converts the whole audio input sentence or word into text . Processing to breakdown the text into smaller understandable pieces which need Machine Learning as a neighborhood. Finally, Data sets of predefined signing are used because the input in order that the software can use AI to display the converted audio into the signing for more development on this track are often done because the ISL dictionary remains small and wishes to grow eventually on process performance.

9. REFERENCE

[1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.

[3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018.

[4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.

[5] Tushar Sonawanev, Shirin Shaikh, Rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.

[6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.

[7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science , 2015.

[8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

[9] Dr .M.Sreedevi, A.Harha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology ,2018.

[10] K.S.N .Murthy, A.V.S.Pavan Kumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.

[11] Deepiika .K, Smitha Vinod, Crime analysis in India using data mining techniques , International journal of Engineering and technology, 2018.

[12] Hitesh Kumar Reddy Toppy Reddy, Bhavana Saini, Ginika Mahajan, Crime Prediction &Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS 2018).