WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

CRIME PREDICTION USING MACHINE LEARNING TECHNIQUES

Dr. Kalyan Bamane[1], Abhishek Gaikwad [1] , Mayank Agarwal[3], Om Naladkar[4], Anisha Lende[5], Pratham Rathod[6]

[1]Associate Professor, Department of Information Technology, D. Y. Patil College of Engineering Akurdi, Pune-44, Maharashtra, India
[2,3,4,5,6]Student, Department of Information Technology, D. Y. Patil College of Engineering Akurdi, Pune-44, Maharashtra, India

ABSTRACT :

Crime prediction is important for preventing crime before it occurs, solving crimes more quickly, and making communities safer. Related work has shown the effectiveness of using machine learning algorithms for crime prediction, but further research is necessary to improve accuracy and efficiency in real- world settings. The proposed system aims to help law enforcement agencies and policymakers in making informed decisions regarding crime prevention and public safety. The platform is designed to utilize predictive algorithms to generate crime incident forecasts. The proposed platform's performance is evaluated using real-world crime data from a selected region, and the results demonstrate that the platform can provide accurate predictions of crime incidents. The Random Forest algorithm outperformed the other algorithms in predicting the crimes with accuracy of 98.07%.

Keywords: Machine Learning, Support Vector Machine (SVM), Random Forest (RF), Decision Tree, Crime, Location.

Introduction :

The prevalence of crime in a society has an effect on its standard of living and level of economic prosperity. It is seen to be a crucial factor in influencing whether people move to a new city or not, as well as which areas should be avoided when travelling. As crime rates rise, law enforcement agencies are still in need of cutting-edge geospatial information systems and cutting-edge data mining approaches to improve crime analytics and better protect their communities. Even though crimes can happen anywhere, it is typical for criminals to concentrate their efforts in the areas that they are most familiar with.
The motivation about the project is that in today's world, all political and governmental bodies worldwide give security a higher priority, with the goal of lowering the occurrence of crime. Since machine learning is the best field to use for applying to high volume crime datasets, and since knowledge acquired from machine learning techniques will be reported to the police force. Through internal investigation and analysis are required to resolve a case based on a specific set of data. The amount of crime data that is now available in Boston makes it difficult for the authorities to evaluate and make decisions on these criminal situations. This study focuses on developing a remedy for the decision-making process of a crime that is committed after identifying them as a key concern.
The use of crime prediction is to analyze and decrease future crime. Using the old data the crime locations are identified. The accuracy, performance, and speed of crime prediction can all be improved with the aid of machine learning techniques. Crime patterns are examined using machine learning approaches using both recent and historical data. The location of the crime can be accurately predicted using crime prediction tools.
By using a machine learning approach to identify the most criminal hotspots and find the place, time, and location, we hope to increase people's awareness of unsafe sites during specific time periods.

Related Work:

Analyzing location-wise crime, predicting crime using machine learning, and developing a web portal for crime analysis and prediction represent a multifaceted approach to addressing contemporary challenges in law enforcement and public safety. Within the realm of crime analysis, understanding the spatial distribution and temporal patterns of criminal activities is essential for effective resource allocation, strategic planning, and proactive crime prevention efforts. Leveraging machine learning techniques for crime prediction offers a promising avenue for identifying high-risk areas and forecasting future criminal incidents based on historical data, demographic factors, socio-economic indicators, and environmental variables. By employing algorithms such as Random Forest, Support Vector Machines (SVM), and neural networks, predictive models can be trained to recognize complex patterns and

relationships within crime data, enabling law enforcement agencies to prioritize resources and deploy interventions more strategically. Moreover, the development of a user-friendly web portal for crime analysis and prediction serves as a valuable tool for stakeholders, including law enforcement officials, policymakers, researchers, and the general public. Such a portal can provide interactive visualizations, real-time updates on crime trends, and access to predictive analytics, empowering users to make informed decisions, collaborate on crime prevention initiatives, and engage with their communities. Through the integration of geospatial data, advanced analytics, and user-centered design principles, the web portal can facilitate data-driven decision-making, enhance situational awareness, and foster greater transparency and accountability in addressing public safety concerns. Overall, the convergence of location-wise crime analysis, machine learning-based prediction, and web portal development represents a holistic approach to leveraging technology and data-driven insights for proactive crime prevention and community-oriented policing strategies in the digital age.

Table 1 - Literature review

Sr.	Reference	Technology Used	Algorithm	Dataset	Limitation
1	A Framework to Predict Social Crime through Twitter Tweets by Using Machine Learning	Python, Jupyter Notebook	Multinomial Naïve Bayes (MNB), KNearest Neighbors (KNN) and Support Vector Machine	Data obtained fromsocial media website	To improve the system's effectiveness and sturdiness, more crime classes can be added.
2	Crime type and occurrence prediction using machine learning algorithm	Python, colab	KNN	Denver city (Kaggle)	They have not implemented machine learning algorithms. The performance Of the algorithm is also calculated by using some standard metrics
3	Predicting Crime Using Time and Location Data	Python	Random Forest, Decision Tree, Bagging and AdaBoost, Extra Trees	Chicago Police Department	They have got least accuracy from classification algorithms
4	Crime prediction and analysis using machine learning	Python, html, css, node.js, Web API	KNN, Naïve bayes, Decision tree	Chicago police department official dataset	Only predicts crime type, not location and time
5	Crime Prediction and Forecasting using Machine Learning Algorithms	PyTorch Framework and Python Libraries	Random Forest, KNN, AdaBoost, Neural Network	Chicago Police dataset	Area considered for prediction was small.
6	Design and Analysis of Machine Learning Algorithms For the reduction of crime rates in India	Time series analysis, Python libraries	LSTM, KNN, CNN, Random Forest	Coastal city of China dataset	Requires high computational power

Problem Statement

The problem statement for the research on "Designing a machine learning-based crime prediction model and developing a user-friendly web application for proactive crime prevention and increasing public safety" revolves around the necessity to create an advanced tool that can predict crime and aid in its prevention. The primary objective is to leverage machine learning techniques to forecast criminal activities, thereby enabling law enforcement agencies and the public to take proactive measures. This research is focused on addressing the limitations of traditional crime prevention methods by introducing a data-driven approach. At its core, this research aims to achieve a dual-fold objective. The first and foremost goal is to enhance the accuracy and reliability of crime prediction. By utilizing machine learning algorithms, the model can analyze vast amounts of historical crime data, identify patterns, and predict potential crime hotspots and trends. Simultaneously, the research places a strong emphasis on the development of a user-friendly web application as a critical aspect of crime prevention. Recognizing the importance of accessibility and ease of use, the proposed solution aims to present complex predictive data in an intuitive and comprehensible manner. The web application will be designed with a user-centric approach, ensuring that both law enforcement professionals and the general public can easily navigate and utilize the tool. Features such as interactive maps, real-time alerts, and customizable dashboards will enhance the user experience, making it more engaging and practical.

4.2 Prediction of crime using machine learning

The second goal is to predict crime using machine learning algorithms. The portal will offer predictive analytics capabilities that allow users to forecast potential crime occurrences based on historical data. This involves identifying factors contributing to crime rates in various areas and providing alerts and recommendations for preventive measures.

4.3 To develop a web portal for crime analysis and prediction

The objective of developing a web portal for crime analysis and prediction is to create a comprehensive online platform that provides tools and features for understanding and anticipating crime patterns. This portal will serve as a central resource for law enforcement agencies, policymakers, and the public to access and analyze crime data effectively. By combining data visualization and machine learning, the portal aims to enhance the ability to monitor crime trends, identify hotspots, and forecast future crime occurrences.

4.4 Ensure Data Security and Privacy:

Implement robust security measures to protect sensitive data used in the predictive model, ensuring privacy and compliance with relevant regulations. Ensure the application operates transparently, providing clear explanations of how predictions are made and fostering trust among users.

Comparison

In the below table it has shown comparison of various algorithm based on various factors with respect to reference, technologies used in that algorithm, datasets, algorithm’s accuracy and also its limitation.

Reference	**Technology Used**	**Algorithm**	**Dataset**	**Accuracy**	**Limitation**
A Framework to Predict Social Crime through Twitter Tweets by Using Machine Learning	Python, Jupyter Notebook	Multinomial Naïve Bayes (MNB), KNearest Neighbors (KNN) and Support Vector Machine	Data obtained fromsocial media website	SVM: 92.0%	To improve the system's effectiveness and sturdiness, more crime classes can be added.
Crime type and occurrence prediction using machine learning algorithm	Python, colab	KNN	Denver city (Kaggle)	93.07%	They have not implemented machine learning algorithms. The performance Of the algorithm is also calculated by using some standard metrics

Predicting Crime Using Time and Location Data	Python	Random Forest, Decision Tree, Bagging and AdaBoost, Extra Trees	Chicago Police Department	Extra Trees: 88.92%	They have got least accuracy from classification algorithms

Crime prediction and analysis using machine learning	Python, html, css, node.js, Web API	KNN, Naïve bayes, Decision tree	Chicago police department official dataset	78.9%	Only predicts crime type, not location and time
Machine learning based analytical Approach for geographical analysis and prediction of Boston City crime using geospatial dataset	Folium, Seaborn and Matplotlib libraries	Random Forest, PCA, Decision Tree	Boston City Dataset	69%	Accuracy of the model was not satisfying.
Crime Prediction and Forecasting using Machine Learning Algorithms	PyTorch Framework and Python Libraries	Random Forest, KNN, AdaBoost, Neural Network	Chicago Police dataset	90.77%	Area considered for prediction was small.
Design and Analysis of Machine Learning Algorithms For the reduction of crime rates in India	Time series analysis, Python libraries	LSTM, KNN, CNN, Random Forest	Coastal city of China dataset	81%	Requires high computational power

According to comparison, research done and literature survey it can be found that Random Forest Algorithm is most efficient in terms of speed, time, throughput and other factors

1.1 History of Random Forest Algorithm:

The random forest algorithm, introduced by Leo Breiman in 2001, represents a significant advancement in ensemble learning. By combining the principles of decision trees with bagging, random forests offer improved predictive accuracy and robustness. Unlike traditional bagging methods, random forests introduce randomness by selecting a random subset of features at each split, thereby reducing the correlation between individual trees and enhancing the ensemble's performance. Breiman's innovation also includes the concept of out-of-bag (OOB) error estimation, allowing for efficient model evaluation without the need for a separate validation set. This approach has made random forests particularly attractive for tasks with high-dimensional data or complex relationships. Over the years, random forests have become a cornerstone in machine learning, applied across diverse domains such as finance, healthcare, and natural language processing. Their versatility, ease of implementation, and ability to handle both classification and regression tasks have contributed to their widespread adoption. Despite the emergence of newer ensemble methods, random forests remain a go-to choice for many practitioners due to their strong performance and interpretability. Ongoing research continues to refine and extend the capabilities of this influential algorithm, ensuring its relevance in the ever-evolving landscape of machine learning and data science.

4.2. Basic Structure of Random Forest Algorithm:

The basic structure of the random forest algorithm encompasses several crucial elements. At its core, it comprises an ensemble of decision trees, each trained on a different bootstrap sample of the data. Random feature selection adds diversity by considering only a subset of features at each node split, thereby reducing correlation among trees. During prediction, the ensemble typically employs a majority voting scheme for classification or averaging for regression. An essential aspect of random forests is the use of out-of-bag samples for model evaluation, eliminating the need for a separate validation set. Hyperparameters like the number of trees and maximum tree depth can be fine-tuned to optimize performance. Moreover, random forests lend themselves well to parallelization, facilitating efficient training on large datasets. This amalgamation of techniques results in a robust and versatile algorithm capable of handling various tasks across domains, making random forests a popular choice in machine learning applications.

4.3 Evaluation criteria for Random Forest algorithm:

Three important criterions were used by NIST to evaluate the algorithms that were submitted by experts.

1. **Security:** Robustness to Adversarial Attacks is used which uses assessing the model's resilience to adversarial manipulation of input data, ensuring that it maintains accuracy and reliability even when faced with malicious attempts to deceive it. Its privacy prevention is done by evaluating the algorithm's ability to handle sensitive data securely, protecting against unauthorized access or information leakage, particularly in scenarios involving personal or confidential information.
2. **Cost Another:** Computational Efficiency is measuring the algorithm's performance in terms of computational resources required for training and inference, ensuring that it can deliver accurate results without excessive processing time or hardware requirements. Scalability here is assessing the algorithm's ability to handle large-scale datasets and computational workloads efficiently, enabling cost-effective deployment in real-world applications with varying data volumes and processing demands.
3. **Algorithm:** Evaluating the algorithm's adaptability to different problem domains and data types, ensuring that it can accommodate diverse requirements and challenges encountered in practical scenarios. Assessing the simplicity of the algorithm's design and implementation, facilitating ease of understanding, deployment, and maintenance for developers and practitioners. Considering the algorithm's compatibility with various hardware architectures and software environments, ensuring seamless integration and optimal performance across different platforms and technologies.

Proposed Methodology

The proposed methodology for analyzing location-wise crime and predicting crime using machine learning, as well as developing a web portal for crime analysis and prediction, involves several key steps. Firstly, comprehensive datasets containing location-wise crime data, including various types of crimes, time stamps, and geographic coordinates, will be collected and preprocessed to ensure data quality. Exploratory data analysis (EDA) will then be conducted to understand the distribution of crime incidents and identify patterns and hotspots of criminal activity. Relevant features, including demographic data, socio-economic indicators, and weather conditions, will be extracted and engineered to enhance Machine learning algorithms have emerged as a crucial tool in predicting and preventing criminal activities. Crime prediction models empower law enforcement agencies and policymakers with insights into crime patterns, enabling proactive measures for public safety. The increasing availability of data coupled with sophisticated algorithms like Random Forest, Gradient Boost, AdaBoost, Support Vector Machines, and K-Means presents a significant opportunity to develop more accurate crime prediction systems. Simultaneously, a web portal will be developed to provide an intuitive interface for accessing and visualizing crime data. Interactive maps, charts, and dashboards will be integrated to display crime statistics, trends, and predictions for different locations. Users will be able to explore historical crime data, receive real-time updates, and generate predictive insights based on machine learning models through the portal. Once developed, the web portal will be deployed on a reliable hosting platform, with regular updates to the crime prediction models using new data to ensure accuracy and relevance. Security measures will be implemented to protect sensitive information and ensure user privacy. Overall, this methodology aims to provide valuable tools for law enforcement agencies and policymakers in crime prevention and management efforts by leveraging data-driven insights and predictive analytics.

1.2 Algorithmic Details

Machine learning is the automatic discovery of significant patterns in data. Over the past few decades, it has evolved into a widely used technique in almost any endeavour that calls for information extraction from enormous data sets. Anti-spam software learns to filter our email communications, search engines learn how to give us the best results (while placing professional table advertising), and software that can detect frauds protects credit card transactions. Smartphones with intelligent personal assistance software can recognise faces in photos and interpret voice instructions. Automobiles are equipped with accident-prevention systems created using machine learning algorithms. We had three different classification issues to address following the preprocessing mentioned in the earlier parts, therefore we began by using an assortment of classification algorithms. The following are the algorithms which we are using :

:***5.5.1. KNN (K- Nearest neighbors):***

K-Nearest Neighbours is a straightforward and well-liked machine learning approach that is utilised for both classification and regression tasks. Given that it is non-parametric, it does not assume anything regarding the distribution of the data. The distance between each new data point and each data point already present in the training dataset is calculated by the KNN algorithm. Based on the estimated distances, it then chooses the K closest neighbours to the new data point. The majority class label of the new data point's K closest neighbours is then applied.

5.5.2 Decision Tree:

As the name implies, it is a tree that supports our ability to make decisions. It is a very fundamental and significant predictive learning method used for both classification and regression. It differs from others since it makes decisions one at a time and operates intuitively. Non-parametric: Quick and effective The dataset is divided depending on the most significant variable by the decision tree using some sophisticated criteria. The goal is to get to the point when homogenous subsets are making predictions with the highest degree of certainty

5.5.2 Random Forests:

A highly well-liked ensemble learning technique called Random Forests creates a lot of classifiers on the training data and then combines all of their results to produce the best predictions on the test data. In order to prevent overfitting on the training data, the Random Forests algorithm incorporates randomization while making split choices. It is a variance-minimizing technique.

Modules

In the pursuit of advancing user authentication methodologies, this research introduces a set of meticulously designed modules to comprise a comprehensive and secure authentication system. Each module plays a distinctive role in fortifying the authentication process, combining innovative techniques and advanced algorithms to ensure a robust yet user-friendly experience. The following section provides detailed introductions to each module, outlining their specific functionalities and contributions to the overall framework.

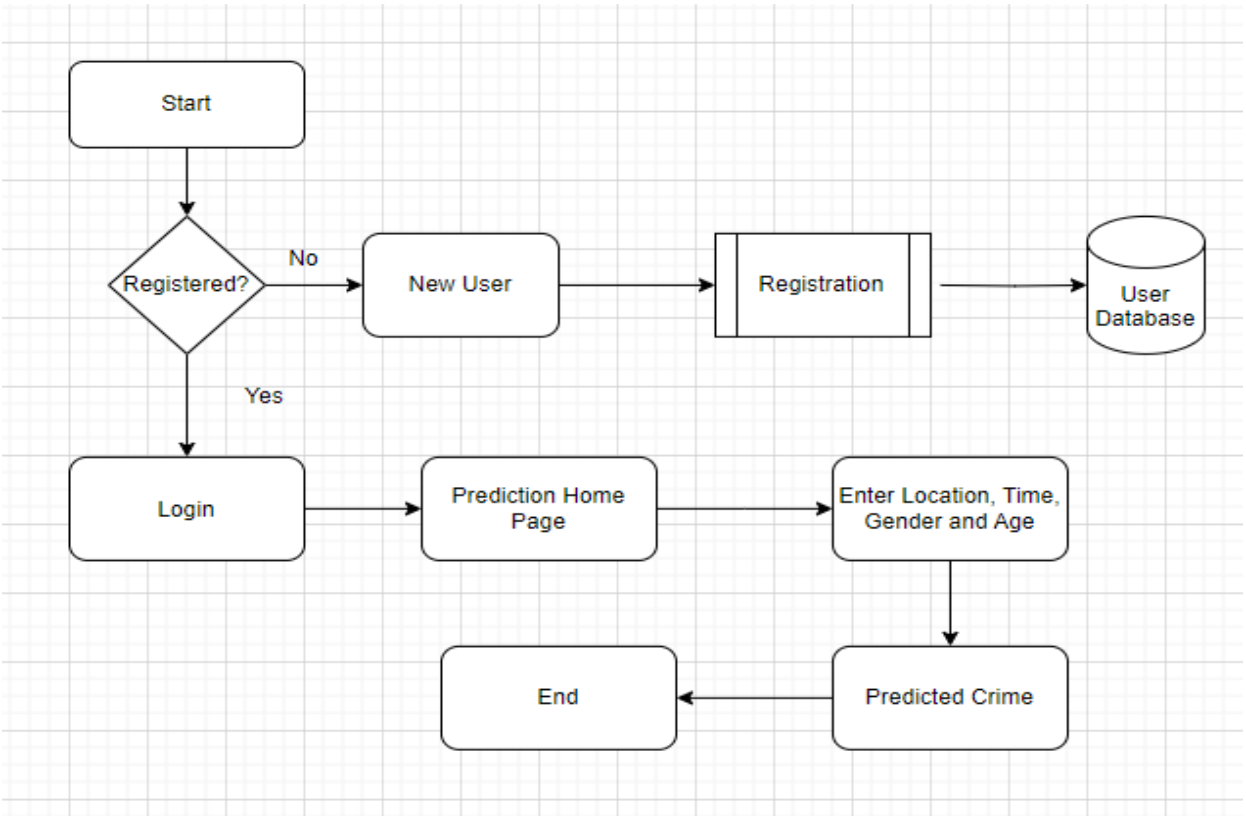

Fig. 2 – System Architecture

1.3 Register Module:

The Register Module constitutes the initial phase in user interaction, focusing on capturing essential details vital for authentication. Users are prompted to provide key information..

Phase	**Registration**
Initial State	Username, password, email, etc.
Final State	Registration success
Algorithm	IF isValid (username, password, email, phone, etc.) and isTrue (key) THEN Registration_success .ELSE Registration_failed

1.4 Login Module:

The Login Module serves as a critical juncture in the authentication workflow, involving a meticulous verification process for the user-provided username and password. Employing a secure comparison mechanism, this module utilizes advanced algorithms to thoroughly validate user credentials.

Phase	**Login**
Initial State	Username, Password
Final State	Username is valid, password is valid, username is invalid, password is invalid
Algorithm	IF isTrue (username, password) and isValid THEN Username_valid ELSE Username_invalid, Password_invalid

1.5 Prediction home page:

Facilitating seamless user interaction. Users are required to employ, adding an extra layer for guiding new user. The module is intricately designed to ensure the navigate user for next steps.

1.6 Upload Datails of Crime :

Users are prompted to enter the details of crime and location which he has to check, fortifying the processing process with a multi-faceted approach to find information.

Phase	**Entering Data**
Initial State	Crime Type, Location,etc
Final State	Entered Data is not valid
Algorithm	IF isValid (Crime) and isTrue (key) THEN send crime type

1.7 Predicting Crime:

Upon successful completion of the entering details, the Successful processing Module serves as the result, granting users access to the cime type. This pivotal stage enables users to navigate the for next action, with a steadfast commitment to maintaining data confidentiality and integrity throughout their interaction with the system.

Phase	**Successs**
Initial State	Crime Type,
Final State	Crime Type.
Algorithm	IF isValid (Crime) and isTrue (key) THEN send crime type

Conclusion

In this project, we proposed a machine learning-based web portal for predicting crime incidents in a given area. The platform utilizes various crime reports data to generate accurate crime incident forecasts. We evaluated the proposed platform's effectiveness using real-world crime data from a selected region and demonstrated that the platform can provide accurate predictions of crime incidents. The platform's potential benefits include preventing crime, increasing public safety, and informing policy makers decisions. The proposed ML web portal offers a promising tool for predicting crime incidents and informing crime prevention strategies. The platform's ability to process vast amounts of data from various sources and generate accurate predictions of crime incidents has the potential to revolutionize the way law enforcement agencies and policymakers tackle crime prevention. Future work could focus on incorporating additional data sources, such as social media data and CCTV footage, to improve the platform's accuracy further. Additionally, the platform's user interface could be enhanced to allow users to interact with the data and explore crime patterns more intuitively. Finally, the platform could be integrated with existing crime reporting systems to provide real-time crime incident forecasts to law enforcement agencies and policymakers.

REFERENCES :

1. Sonal Bawankule, Ajay Bokde, Akhilesh Bolgundewar, Simran Patil, Minar Patil, “CRIME RATE PREDICTION SYSTEM USING MACHINE LEARNING”, International Research Journal of Modernization in Engineering Technology and Science” 2022.
2. Vishan Kumar Gupta, Surendra Kumar Shukla, Anupriya, Ramesh Singh Rawat “Crime Tracking System and People's Safety in India using Machine Learning Approaches” International Journal of Modern Research, Vol. 2, No. 1, 1-7, 2022.

3. S.K. Senthil Kumar, Gampa Adarsh, Jessu Shashank, Md. Abdul Sameer, "Crime Prediction and Analysis Using Machine Learning", International Journal Of Tecno- Engineering Volume Xii, Issue Ii,2021.
4. Kanimozhi N, Keerthana N V, Pavithra G S, Ranjitha G,Yuvarani S "CRIME TYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING ALGORITHM" International Conference on Artificial Intelligence and Smart Systems(ICAIS- 2021) IEEE
5. Krishnendu S.G, Lakshmi P.P, Nitha L "Crime Analysis and Prediction using Optimized K-Means Algorithm" International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE..
6. Atharva Deshmukh, Saurabh Banka, Sean Bruno Dcruz, Sana Shaikh, Amiya Kumar Tripathy, "Safety App: Crime Prediction Using GIS", 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA).
7. Sunday Adeola Ajagbe, Roseline Idowu, John Babalola Oladosu, Ademola O. Adesina,"Accuracy of Machine Learning Models for Mortality Rate Prediction in a Crime Dataset", International Journal of Information Processing and Communication (IJIPC)Vol. 10 No. 1&2 (December, 2020).
8. Shuyu Yao, Ming Wei, Lingyu Yan, Chunzhi Wang, Xinhua Dong, Fangrui Liu, YingXiong, "Prediction of Crime Hotspots based on Spatial Factors of Random Forest", The15th International Conference on Computer Science & Education (ICCSE 2020) August 18- 20, 2020.
9. Krishnendu S.G, Lakshmi P.P, Nitha L "Crime Analysis and Prediction using Optimized K-Means Algorithm" International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE.
10. Ch. Mahendra, G. Nani Babu, G. Balu Nitin Chandra , A. Avinash, Y. Aditya, "CRIME RATE PREDICTION", Journal Of Engineering Sciences(JES) 5 May, 2020.