Analytics of Crime Against Women Using ML Algorithms

[1]Annalaxmi Valluvar, [2]Spoorti Shetty, [3]Subhashree Pandian, [4]Prof. Suvarna Chaure

[1,2,3]T.E. Student, Dept. of Computer Engineering, SIES GST, Navi Mumbai, India.

[1]annalaxmivalluvarce120@siesgst.ac.in, [2]spoortishettyce120@siesgst.ac.in,

[3]subhashreepandiance120@siesgst.ac.in

[4]Assistant Professor, Dept. of Computer Engineering, SIES GST, Navi Mumbai, India.

suvarnac@sies.edu.in

Abstract: Everybody understands that crimes against women rank among the most worrying problems for society and several law enforcement agencies worldwide. Data analytics has recently emerged as a popular method for evaluating data, extracting its information, and relating it in a variety of application fields. The analysis of these data enables us to keep track of events, spot patterns more efficiently between incidents, find resources, and take quicker decisions as needed. Machine learning, a cutting-edge and expanding field of study that may develop patterns and procedures across numerous sectors to draw conclusions about usable information, is one of the most crucial technologies. The police department can focus its resources there by using our technology to identify the crime type that is most likely to occur more extensively on the provided data collection of crimes against women. Data mining is employed in this study to glean new, useful information from the provided data set. The proposed system uses its dataset and the algorithm created to predict the precise sort of crime that will take place next. This forecast lessens the complexity of the time frame and helps the police force solve the crime.

Keywords — SVM, KNN, EDA, Gradio

I. INTRODUCTION

Crime analysis is a function of law enforcement that necessitates methodical investigation in order to uncover patterns and trends in crime and disorder. These data on patterns and trends can aid detectives in locating and apprehending criminals and help law enforcement agencies locate resources more effectively. Crime analysis is important for developing crime prevention strategies and for solving crime-related issues. Crime analysis is mostly used to support police agency operations. These duties encompass criminal investigation, comprehension, and prosecution as well as planning for reducing crime, problem-solving, and the evaluation and accountability of police work.

The science of machine learning is the use of computers to make choices without human intervention. Recently, machine learning has been used in web search, speech recognition, self-driving automobiles, etc. Additionally, using cited data to forecast crime is now possible thanks to ML. Machine learning-based crime analysis often entails data gathering, classification, pattern recognition, prediction, and visualization.

Any nation, state, or district's biggest problem is crime against women. To handle the problem, it is necessary to gather timely and relevant information. Investigating and identifying crime as well as the connections between different criminals is crime analysis. Numerous analytical or predictive data mining techniques have been developed and are applied in diverse fields. Many researchers are employing various data mining tools to deter and manage various crimes. The creation of an accurate prediction model for the crime is the main goal of this effort. In order to evaluate the Indian crime dataset generated between 2001 and 2014, classification techniques, namely Random Forest, SVM (Support Vector Classifier), KNN (K-Nearest Neighbor) are used in the work.

The rest of this paper is organized as follows: Firstly, Chapter 2 provides a detailed assessment of the literature. The project's methodology in which various algorithms that are employed for project development, and the results that are obtained, are all addressed in Chapter 3. The project's conclusion and the project's future support are then discussed in the paper's conclusion i.e., in Chapter 4.

II. LITERATURE REVIEW

This section gives an overview of the approaches used in the literature for crime prediction, using various ML algorithms. In the recent decade, several strategies have been proposed, but each has significant differences in terms of algorithms, categorization, and accuracy. With this preliminary survey, we could devise a strategy for carrying

 DOI : 10.35291/2454-9150.2023.0049

out the project. Below is a summary of the many methodologies and categorization strategies for yoga position estimate.

The study [1] discusses several ML algorithms for crime type and occurrence prediction. The main goal of the study is to determine which kind of crime contributes the most, along with the frequency and location of the crimes. In comparison to pre-composed works, the Naïve Bayes method suggested in the work produced results with a rather high accuracy.

KNN, Decision trees, and other crime prediction algorithms are discussed in the work Crime Prediction and Analysis [2]. The main objective of the work is to show the effectiveness and worth of machine learning in predicting violent crimes occurring in a certain location so that law enforcement may use it to reduce crime rates in society. According to the findings, Decision tree, KNN, and Extra tree classifiers produced the greatest results in terms of training efficiency and accuracy.

Crime Analysis Mapping, Intrusion Detection - Using Data Mining [3], discusses Crime mapping analysis based on KNN (K – Nearest Neighbor) and ANN (Artificial Neural Network) algorithms to streamline the process and find the areas where crimes occur the most frequently. With the aid of the SAM tools, the work avoids a difference in the results, and after that, the subsequent information will be used to discover relationships between those.

The purpose of the work, Criminal Behavior Analysis and Segmentation using K-Means Clustering [4], is to comprehend the idea of data mining and machine learning, which can be utilized to uncover criminal patterns and behaviors. The primary goal of the article is to provide a brief overview of how machine learning may be used by the legal system to recognize, foresee, and highlight infractions at a much faster rate.

Using historical public property crime data from 2015 to 2018 from an area of a large coastal city in southeast China, the research article Comparison of Machine Learning systems for Predicting Crime Hotspots [5] evaluates the accuracy of several ML systems. Using historical crime data as a basis, the LSTM model fared better on implementation than KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks.

In the research paper Diagnosis of Crime Rate Against Women using K-fold Cross Validation through Machine Learning Algorithms [6], it is discussed how six different machine learning techniques, such as KNN and decision trees, Naive Bayes, Linear Regression CART (Classification and Regression Tree), and SVM, are used to diagnose the crime rate versus women. According to the findings, the KNN algorithm performs more effectively than other ML algorithms. In this essay, the most frequently occurring crime types and regions in India were discussed.

In the paper [7], the most significant aspects of crime are retrieved using univariate and bivariate exploratory analysis. In order to pinpoint the most common criminal reasons, the article will look at several statistical data. Using the Akaike Information Criteria (AIC) method, unimportant features are removed. The model is then assessed using the Mean Absolute Error (MAE), Median Squared Error (MSE), and Root Mean Squared Error (RMSE) methods.

III. PROPOSED SYSTEM

The proposed system of analytics for crime against women will be a data-driven approach that leverages ML algorithms to analyses and understand crime patterns, and predict incident rate. A thorough dataset that contains details on criminal episodes and the situation where they happened will serve as the foundation for the system.

The system identifies patterns and trends in the data by combining supervised and unsupervised learning techniques. To categorize different sorts of crimes and forecast the likelihood (high/low) of future occurrences, supervised learning techniques like decision trees, random forests is employed. To find patterns and group similar instances together, unsupervised learning methods like k-means clustering are implemented.

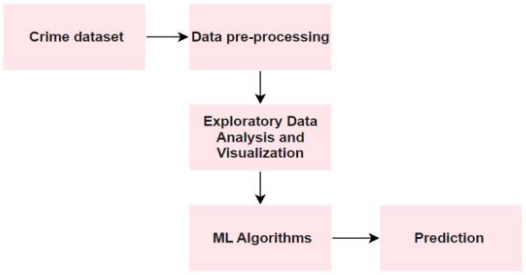


Figure 1. shows the model which contains 6 stages of how the system works.

Figure 1: Crime Prediction System

IV. METHODOLOGY

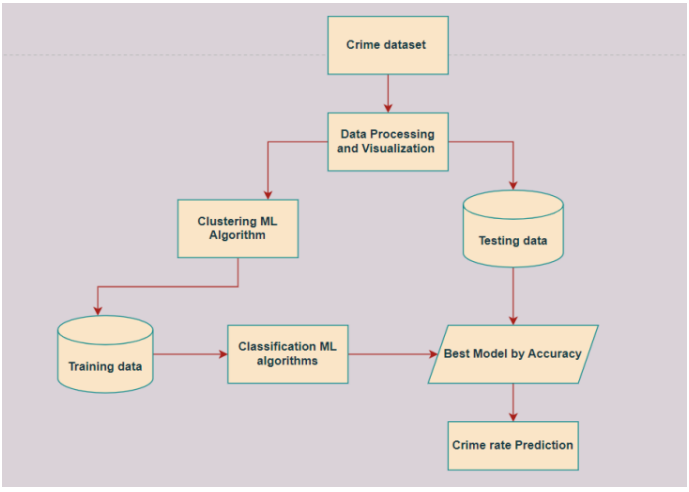


Figure 2: Workflow

In Figure 2., we have discussed the flow and working of

 DOI : 10.35291/2454-9150.2023.0049

themodel.

A. *Crime dataset*

From the Kaggle website, the dataset for crime against women is chosen. The selected Indian dataset spans the years 2001 to 2014. The record contains information on the state and district where the crime occurred, as well as information about the crime's type, such as rape, kidnapping, dowry deaths, Assault on women with intent to outrage her modesty, Insult to modesty of Women, Cruelty by Husband or his Relatives, Importation of Girls.

B. *Data pre-processing*

This method removes main errors and contradictions that are expected when multiple resources of data are getting into the dataset. In the project, missing data are handled which is the most important step in data cleaning. The next primary step is handling values from mixed data sets i.e., for example, if in the dataset we have UP and Uttar Pradesh which means the same, are handled by making them as one set by naming it either UP or Uttar Pradesh.

C. *Exploratory Data Analysis (EDA)*

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

In EDA, all possibilities are discussed with are as follows:

Figure 3 shows the Year wise analysis of the whole dataset, analyzing a dataset year-wise can help identify trends and patterns over time and use this information to make predictions or develop insights about future trends.

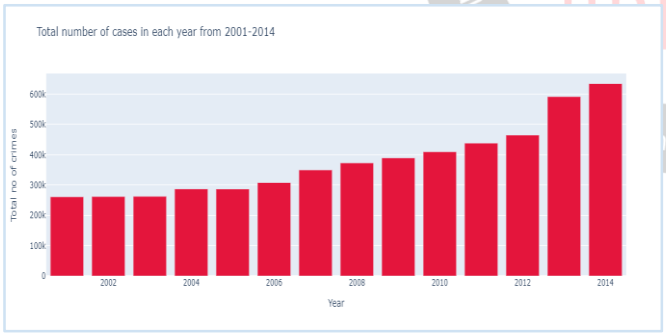


Figure 3: Year wise crime analysis

Figure 4 shows that the total cases in each year from the dataset that ranges from 2001 – 2014. As the figure, the bar of cases is increasing year by year.

Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives
2001	32150	29290	13702	68248	19492	98340
2002	32746	29012	13644	67886	20310	98474
2003	31694	26592	12416	65878	24650	101406
2004	36466	31156	14052	69134	20002	116242
2005	36718	31500	13574	68350	19968	116638
2006	38696	34828	15236	73234	19932	126256
2007	41474	40832	16186	77468	21900	151860
2008	42934	45878	16344	80826	24428	162688
2009	42794	51482	16766	77422	22018	179092
2010	44344	59590	16782	81226	19922	188082
2011	48412	71130	17236	85936	17140	198270
2012	49846	76524	16466	90702	18346	213054
2013	67414	103762	16166	141478	25178	237732
2014	73470	114622	16910	164470	19470	245754

Figure 4: Total number of cases in each year from 2001 -2014

	Count
Cruelty by Husband or his Relatives	2233888
Assault on women with intent to outrage her modesty	1212258
Kidnapping and Abduction	746198
Rape	619158
Insult to modesty of Women	292756
Dowry Deaths	215480

Figure 5 shows all the crimes types by 2014 in two forms: Highest and Lowest

Figure 5: Highest vs Lowest reported crime by 2014

In Figure 6, the proportion of each crime category from 2001 to 2014 is depicted in a pie chart. Pie charts with percentages may be used to visualize the relative frequency of various crime categories in each dataset. The percentages may be used to quickly summarize the distribution of crime categories as well as to highlight the most prevalent sorts of crimes.

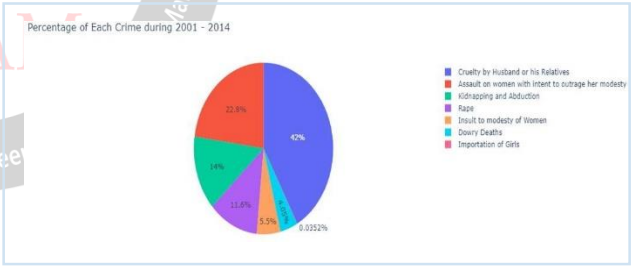


Figure 6: Representation of percentage of each crime during 2001 - 2014

Figure 7 to figure 12 represents year wise analysis of all crime types. Year-wise analysis of all crime types is important because it can help identify patterns and trends in criminal activity over time. By analyzing crime data over a period of several years, ML algorithms can identify patterns in crime rates etc.

 DOI : 10.35291/2454-9150.2023.0049

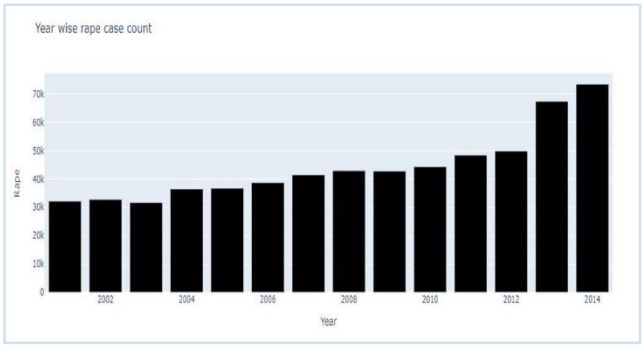


Figure 7: Year wise rape count

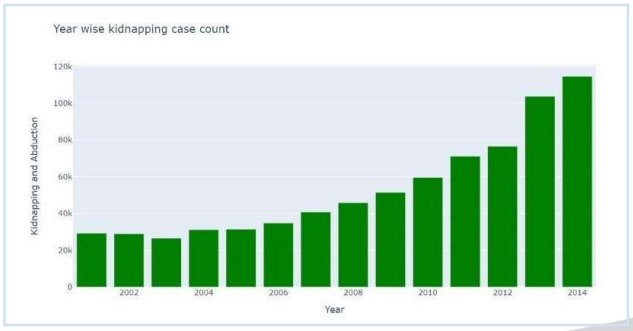


Figure 8: Year wise kidnaping count

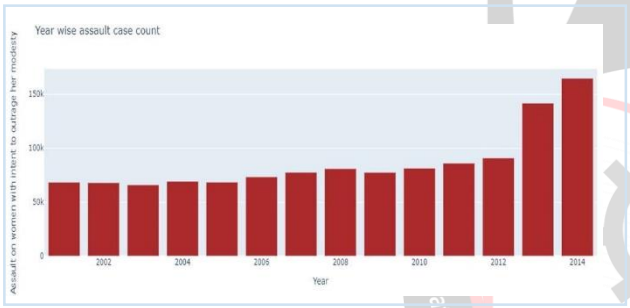


Figure 9: Year wise dowry death case count

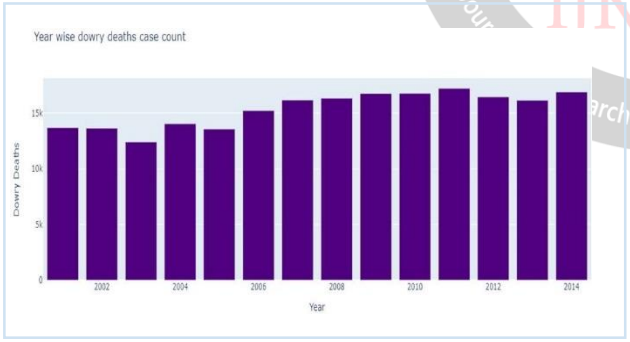


Figure 10: Year wise assault count

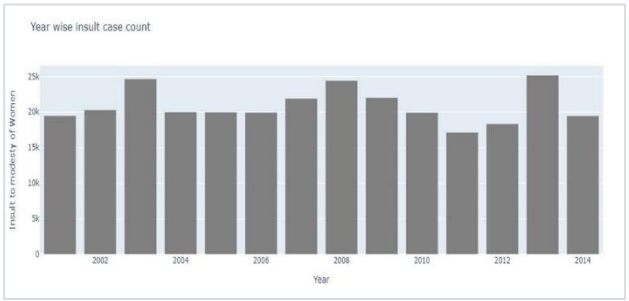


Figure 11: Year wise insult case count

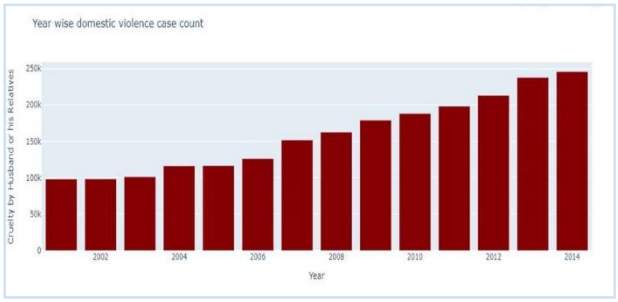


Figure 12: Year wise domestic violence case count

Figure 13 shows all the total crime types distribution according to the states and union territories.

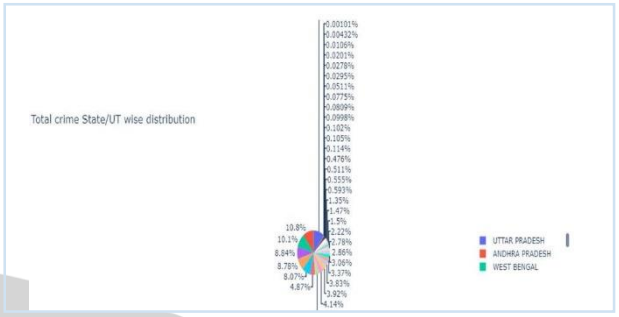


Figure 13: State/ UT wise distribution

Figure 14 shows the comparison of crime rate of year 2001 & 2014

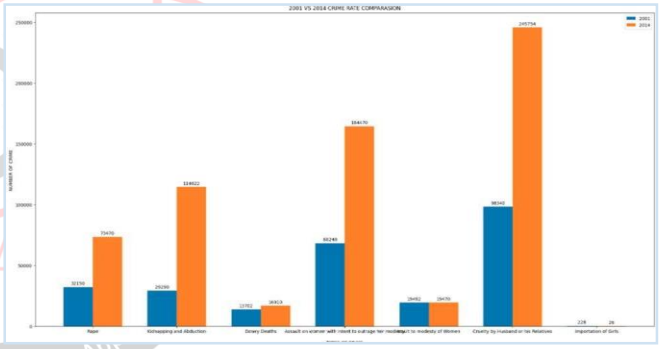
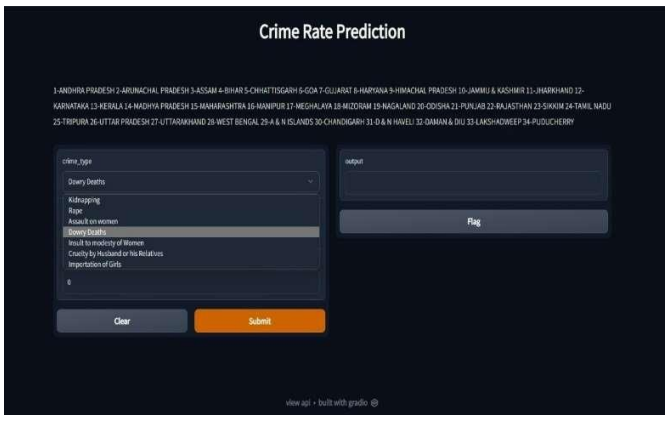


Figure 14: 2001 vs 2014 Crime rate comparison

The UI Design

Figure 15 describes the UI of the work where '0' and '1' defines High crime and low crime rate respectively.



 DOI : 10.35291/2454-9150.2023.0049

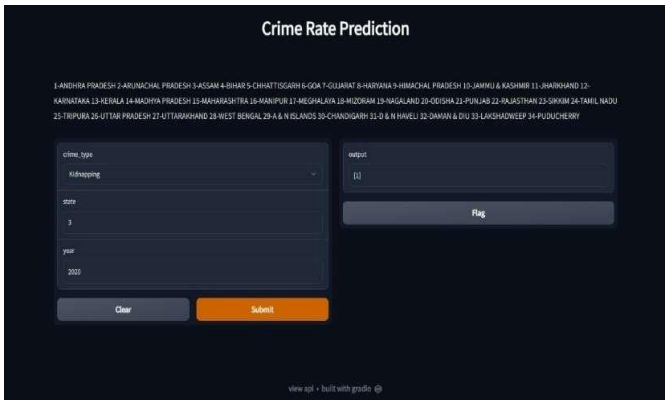


Figure 15: UI design of crime prediction system

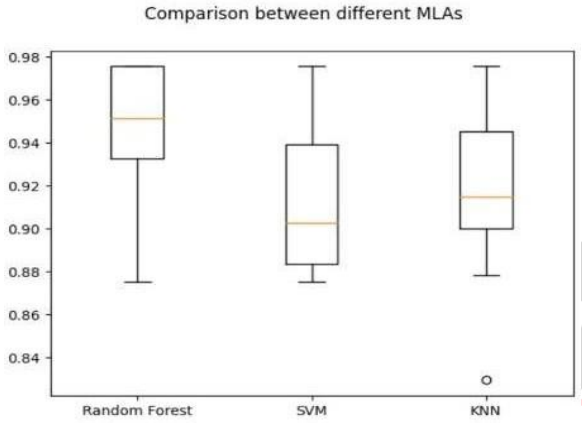


Figure 16: Comparison of ML algorithms based on accuracy

Figure 16 shows the comparison of the algorithms used in the work using box-plot. Box plots are a useful visualization tool for this task, as they provide a clear and concise summary of the distribution of a dataset.

D. Machine Learning Algorithms

In order to tackle the problem of crime trends, explored several state-of-the-art machine learning algorithms. ML techniques like clustering algorithms and classification algorithms are applied. In Clustering algorithm, K-means clustering is used (which is an unsupervised learning method for clustering data points. The algorithm iteratively divides data points into K clusters by minimizing the variance in each cluster) with Classification algorithms like Random Forest, Support Vector Classifier (SVM) and K-nearest neighbor (KNN).

E. Prediction

Machine learning model predictions allow businesses to make highly accurate guesses as to the likely outcomes of a question based on historical data. Using the prior mentioned methods, crime rates are forecasted, and the data accuracy is reached.

V. RESULTS

Figures 3 to 14 display the Exploratory Data Analysis (EDA) of year-by-year crime analysis, study of specific crime types, distribution of cases by State/UT, and comparisons between crime rates.

The user interface (UI) design of the crime prediction system is pictured in Figure 15. Based on the input parameters of crime type, state, and year, it forecasts the likelihood of crime as either high or low.

Figure 16 compares the ML algorithms used in the project based on accuracy data represented as a box-plot.

VI. CONCLUSION AND FUTURE SCOPE

This study exemplifies the potential for crime prediction by deploying machine learning (ML) strategies in a Jupyter notebook with Python as the primary language. The program's findings can be used to determine the predominant types of crimes committed in each state as well as the overall numbers of crimes committed in each of India's states and territories. This would make it easier for the government and law enforcement to pass stricter legislation that will reduce crime rates and make it safer for women to live in society. The techniques used in this research include exploratory data analysis and machine learning (ML) algorithms like K- means, Random Forest, SVM, and KNN, with Random Forest exhibiting the highest accuracy rate of ML algorithms.

In the future, we can continue working on this project using several cross-validation techniques to boost accuracy, and we can also enhance the accuracy of the data pre-processing step. The building of ML models to predict crime hotspots can be done with PySpark. The work's UI design can be upgraded with a variety of cutting-edge trends to make it more appealing and descriptive.

Furthermore, contemporary datasets that are more realistic can be designed to evaluate the scalability and effectiveness of multiple systems.

REFERENCES

[1]Kanimozhi, N., Keerthana, N. V., Pavithra, G. S., Ranjitha, G., & Yuvarani, S. (2021, March). Crime type and occurrence prediction using machine learning algorithm. In *2021 International conference on artificial intelligence and smart systems (ICAIS)* (pp. 266-273). IEEE.

[2] Kumari, Pratibha & Gahalot, Akanksha & Uprant, & Dhiman, Suraina & Chouhan, Lokesh. (2020). Crime Prediction and Analysis. 1-6. 10.1109/IDEA49133.2020.9170731.

[3] Panja, B., Meharia, P., & Mannem, K. (2020, June). Crime Analysis Mapping, Intrusion Detection-Using Data Mining. In *2020 IEEE Technology & Engineering Management Conference (TEMSCON)* (pp. 1-5). IEEE.

[4] Jha, G., Ahuja, L., & Rana, A. (2020, June). Criminal

 DOI : 10.35291/2454-9150.2023.0049

behaviour analysis and segmentation using K-means clustering. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 1356-1360). IEEE.

[5] Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. IEEE Access, 8, 181302-181310.

[6] Tamilarasi, P., & Rani, R. U. (2020, March). Diagnosis of crime rate against women using k-fold cross validation through machine learning. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1034-1038). IEEE.

[7] Shukla, A., Katal, A., Raghuvanshi, S., & Sharma, S. (2021, June). Criminal Combat: Crime Analysis and Prediction Using Machine Learning. In 2021 International Conference on Intelligent Technologies (CONIT) (pp. 1-5). IEEE.

 DOI : 10.35291/2454-9150.2023.0049