DOI : https://doi.org/10.56452/7-5-199

ISSN: 0974-5823 Vol. 7 No. 5 May, 2022

International Journal of Mechanical Engineering

Crime Prediction using Machine Learning and Deep Learning Techniques

Charu Negi

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

ABSTRACT:

Crime is nothing but serious, violent, intentional, unintentional, with reason or without reason rapid action taken by any human. Crime can be considering most dangerous act that can harm victim as well as criminal life as well. Therefore, we are here implementing prediction system using machine learning and deep leaning technology. Goal of this project to use crime dataset for analysis using K-Nearest neighbor classifier with techniques reference to support vector machine. A more accurate algorithm will be utilized for police training once the K-Nearest Neighbor classification and a number of other algorithms are tested for crime prediction. dataset visualization utilizing a variety of graphs. Charts will be created using graphical representations of numerous instances to show, for instance, when criminal activity rates are highest or when the most unlawful crimes occur. This project's entire focus is on how law enforcement may utilize machine learning to identify, anticipate, and solve crimes more quickly, which will lower the crime rate. It is not limited to a certain location since it uses different datasets from different nations or regions and is usable in other states or countries where the dataset is available.

1. INTRODUCTION:

The use of artificial intelligence, and more especially machine learning, provides opportunities to enhance law enforcement's awareness of ongoing and potential criminal conduct while also streamlining the decision-making process for non-criminal matters. The methods of machine learning work well with exceedingly big data sets. Neural networks, often known as NNs, are a type of machine learning approach that is modeled after the way the human brain works and how it interprets neuro-cognitive data. NNs provide high-quality data to enhance police decision-making, and they are the most precise and accurate way of clustering that is currently in use.

This research intends to predict criminal behavior by making use of the myriad of characteristics included in the dataset. The dataset was compiled using information obtained from the relevant official websites. It is possible to apply machine learning algorithms, the primary language of which is Python, to make predictions on the type of criminal activity that will take place in a specific location. The goal would be to train a model such that it could make predictions. The training that was done with the training dataset will be checked with the help of the test dataset. Depending on the level of precision required, a more suitable method will be used to generate the model. The prediction of criminal activity will make use of the k-nearest-neighbor (KNN) classification as well as other techniques. Visual representations of

the dataset are provided for the aim of conducting an investigation into possible offenses that were committed within the nation.

The work is successful in accomplishing both of its goals. The first step is to educate law enforcement groups on the possibilities of NNs and the breadth of research that has been done previously and is currently being done. Academics and developers who are working on automating decision making and planning for law enforcement can benefit from conducting a literature review because it makes it easier for them to share knowledge. The second goal is to demonstrate, through original research, how neural networks (NNs) may be utilized for crime analysis, which may help law enforcement personnel improve their ability to both anticipate and respond to criminal activity. The prior literature assessment demonstrates how the current study is concentrated on predicting increases in either overall criminal activity or the rise in activity for a particular crime. The offered study will advance the greater time horizons of previous research in that it will allow for the prediction of a specific type of criminal activity once the location and time have been determined. When a crime type as well as the time of day are known, another NN model will demonstrate that, in contrast to earlier research that makes use of much broader time horizons, a particular neighborhood in the city may be promptly recognized as the site where the crime is most likely to have occurred. This may be done by determining the location where the crime was most likely to have occurred. According to the findings of the study, neural networks are capable of conducting a diverse array of criminal investigations; nevertheless, they should only be used in advising capacities or for secondary confirmatory analysis. The process of building a model that is capable of making predictions is called predictive modelling. In the technique, a machine learning algorithm is utilized to make those predictions by learning specific features from a training dataset. The training dataset can be thought of as a classroom. Two of the subfields that fall under the umbrella of predictive modeling are regression and pattern classification. Regression models are constructed through the analysis of relationships between variables and trends in order to make predictions regarding the values of continuous variables.

In contrast to regression models, pattern classification aims to assign discrete class labels as an output of a prediction to a particular data value. Predicting whether a day will be sunny, rainy, or snowy may be a pattern classification issue in weather forecasting. A categorization model like this one is an example. Pattern classification tasks may be divided into two categories: supervised learning and unsupervised learning. In supervised learning, the class labels for the input dataset of the classification model are present. By knowing which training dataset had the precise output that will be used to train, we may predict results for unobserved data in a supervised learning scenario.

2. LITERATURE SURVEY

Tamir Azwad, et al [1] have used the methods K-Nearest Neighbors, AdaBoost, Random Forest, and Neural Network to build models to predict crimes from future and their locations. Additionally, they used the various API to visualize the data for data breakdown. The machine learning model using neural networks outperformed the other techniques, obtaining a remarkable accuracy of 90.77%. Additionally, compared to other machine learning models, it demonstrated superior performance w.r.t various precision factor. To find the ideal structure for this type of dataset, they also created a variety of alternative neural network models with various architectures and layer counts. This approach will aid police departments in major urban areas in identifying which calls for service require immediate attention and allocating resources accordingly. Additionally, it would assist them in determining which regions of the city might demand increased police presence in the future.

Shah Neil et al [2] In order to create a system that is much more helpful to law enforcement, the [research] proposed a framework for how computer vision, machine learning, and deep learning may collaborate. The technology in our proposed system may be used to identify persons based on voice notes and locate criminal hotspots. The main hurdle will be creating the system, then there will be problems with its usage and implementation, among other things. However, all of these problems are addressable, and a security system that constantly monitors the whole city may help. To put it another way, picture a situation where they integrate such technology into a police force, obtaining far more reliable tips or leads, and maybe eliminating crime much more quickly.

Only seven crime clusters are considered by Walczak Steven et al [3] NN models can correctly forecast a specific crime cluster 16.4% of the time out of a total of 27 crime clusters, allowing researchers to focus more on the most frequent crimes. This prediction accuracy is clearly better than pure guessing at random, which would provide an estimated prediction accuracy of 14.3% and a value of p0.001 (standard Z-Test). The ability of police to respond could be enhanced by this sort of timely information on the nature of a potential crime. When the exact site of a crime is unknown, a separate NN model locates the incident inside a neighboring zip code 31.2% of the time, according to their analysis. This result shows that, at the very least, certain crime categories are geographically clustered. These NN models show an improvement over the current NN research, which is more limited to predicting general crime over a much longer time horizon. Future study is necessary to further investigate the usage of NNs in crime prediction in order to enhance near-term (immediate) crime predictions. This research may also use extra time cues or various size geographic markers.

Prof. Shivaprasad More et al [4] model helps forecast crime. The perpetrator's age, sex, and relationship may be predicted using a machine learning method. This classifier and regression offer accuracy of around 80%. The dataset may be enhanced and utilized in other countries if the situation is almost same there. The model offers a general forecast of any crime. It's possible to enhance this model using deep learning techniques.

Alkesh Bharati et al. [5] have created a model employing the machine learning concept using a training set of data that has undergone data cleaning and data transformation. The model is able to determine the kind of crime with an accuracy of 0.789. Data visualization makes it simpler to analyze a data collection. Bar, pie, line, and scatter diagrams are among the graphs, and each has distinct qualities. They have developed various graphs and uncovered fascinating statistics in order to better understand the Chicago crime datasets that may aid in capturing the elements that can assist maintain society safe.

Rani A et al [6] studied time series clustering has become increasingly popular recently, and it has been particularly beneficial for identifying practical related trends in crime. The time series clustering technique offers the capacity to analyse a big amount of data. For determining and using data mining techniques, multivariate time series data from a variety of fields, including banking, the health sector, environmental research, and crime, are highly helpful. Law enforcement authorities are interested in learning more about crime trends, which can be discovered by analysing multivariate time series data over time. The trend analysis is used by police administration at the state and district levels to resolve new criminal cases and aid in the prevention of similar crimes in the future. In order to examine crime trends and forecast future crime, this research introduces a novel method that uses records and statistics in combination with the dynamic temporal warping technique and Mahanolobis distance model.

Kim S et al [7] The use of machine learning to predict crimes is the subject of study. In this research, crime statistics from Vancouver for the preceding 15 years are analyzed using two different data-processing approaches. Machine-learning prediction models K-nearest-neighbor

and boosted decision tree are employed to forecast crime in Vancouver, and an accuracy range of 39% to 44% is achieved.

As per the theory of Duan L et al [8] crime has demonstrated the intricate relationships between time, geography, and surroundings as a persistent worldwide issue. The extraction of useful features that might show these complex relationships and forecast where and when crimes will occur is both a hot topic and a research bottleneck. In an effort to use deep convolutional neural networks (CNNs) for automatically extracting features related to crimes, we consequently created the Spatiotemporal Crime Network (STCN). From the retroactive volume of high-dimension data, this model can predict the crime risk for each place in the urban area for the following day. Using crime and many records in US New York City from the year 2010 to 2015, we assessed the STCN. The outcomes demonstrated that STCN obtained 88% and 92% on F1 and AUC, respectively. This demonstrated that STCN's performances were higher than those of four baselines. Last but not least, the projected results were visualised to aid individuals in understanding how they relate to the reality.

Zhao, X et al [9] offer a novel framework called STF that uses transfer learning to predict crime while capturing temporal-spatial trends. STF uses cross-domain urban datasets, such as information on public safety, weather, points of interest (POIs), human mobility, and complaints. There are a number of intriguing research topics, including -Where: determining the Hotpoint for a particular type of crime, like robbery, - When: determining the best or worst times of day to commit a given crime, - Who: identifying potential criminals based on user profiles on online social networks, etc. -How: defining the steps involved in committing an offence; -Why: examining the reasons why an offence occurs at a particular time and in a particular location.

3. PROPOSED SYSTEM:

A. System Architecture

To achieve the ambitious goal of creating a crime prediction system using machine learning and deep learning techniques, the following steps will be taken:

Data Collection: The initial step involves gathering a comprehensive and diverse dataset of historical crime records, socio-economic factors, and other relevant information. This data should include details such as crime type, location, time, and any additional context that may provide insights into the factors contributing to criminal activities.

Data Pre-processing(ETL): The collected data will go thorough ETL pre-processing to check its quality, reliability, and usability for training the machine learning models. This stage will involve cleaning and transforming the data, filling in missing values, and normalizing the features for optimal model performance.

Feature Engineering: The next step is to identify and extract essential features that can provide meaningful insights into crime prediction. This may include, but is not limited to, demographic information, geographic details, economic factors, and temporal patterns. Feature engineering techniques will be employed to transform raw data into a format suitable for training machine learning algorithms.

Model Selection and Development: Various machine learning and deep learning models will be explored to find the most suitable one for predicting crime. Supervised learning techniques such as classification and regression, as well as unsupervised methods like clustering and anomaly detection, will be considered. The chosen model will be further fine-tuned and optimized to maximize its predictive capabilities.

DOI : https://doi.org/10.56452/7-5-199

Model Evaluation: Rigorous evaluation of the developed model will be carried out using appropriate performance metrics such as accuracy, precision and recall. Cross-validation techniques will be employed to ensure the model's robustness and prevent overfitting.

Deployment and Integration: Once the model has been developed and validated, it will be deployed to a suitable platform, enabling real-time crime prediction. This may involve integrating the system with existing law enforcement databases and communication channels to provide authorities with actionable intelligence and facilitate data-driven decision-making.

Continuous Improvement: The system's performance will be regularly monitored and evaluated to identify areas for improvement. As new data becomes available, the model will be retrained and fine-tuned to maintain its accuracy and effectiveness in predicting crime. Feedback from law enforcement agencies and local authorities will be utilized to refine the system further and better address the evolving challenges of crime prevention.

The Communities and Crime Unnormalized Dataset's patterns for violent crime were compared to actual crime statistics data. Three algorithms—KNN, SVM, and CNN—were applied to datasets from communities and actual criminal activity. The test participants were selected at random. The goal of the research was to show the efficacy and accuracy of machine learning (ML) algorithms in predicting violent crime patterns and other applications, including locating criminal hotspots, creating criminal profiles, and spotting criminal trends.

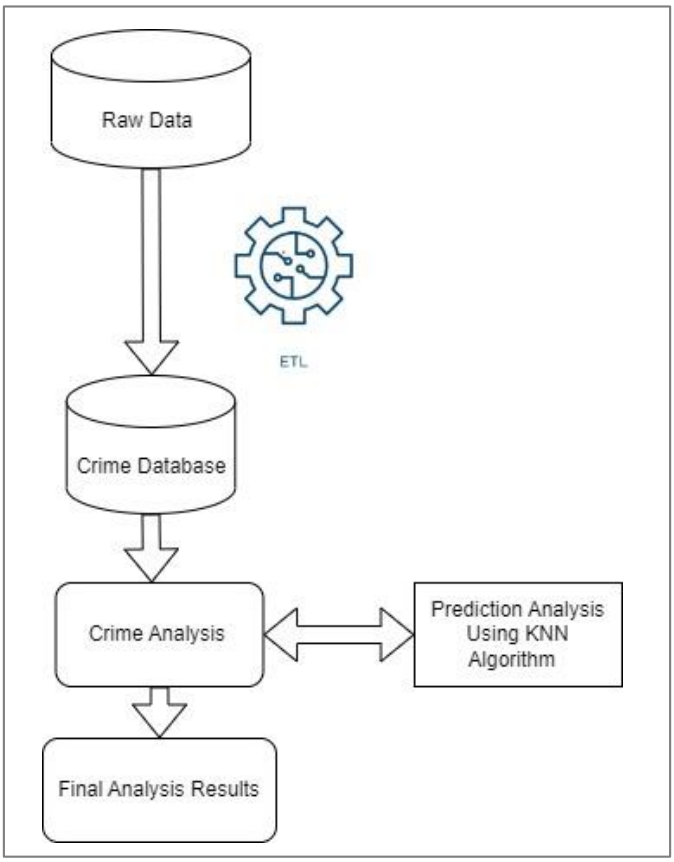

Figure 2: Crime Analysis and Prediction using ML Classifier Algorithm

B. Algorithm: KNN Classifier Algorithm:

We are using KNN (K-Nearest Neighbor) Algorithm as below:

DOI : https://doi.org/10.56452/7-5-199

This method is used to overcome the classification model issues. In essence, the K-nearest neighbor approach, sometimes referred to as K-NN, creates artificial borders to categorize the data. As new data is received, the algorithm will make an effort to predict it to the closest boundary line.

Larger k values provide smoother separation curves, which result in less complex models. On the other side, models with lower k values tend to overfit the data and be more complex. The correct k-value must be used while analysing the data to avoid the dataset being overfitted or underfitted.

Using the k-nearest neighbor method, we fit the historical data (or train the model) and then extrapolate the future. Sample: Let us consider two points, A1(X1, Y1) and B2(X2, Y2). Then the Euclidian distance of the two points as the following

$$\text{Distance} = \sqrt{[(x2 - x1)^2 + (y2 - y1)^2]}$$

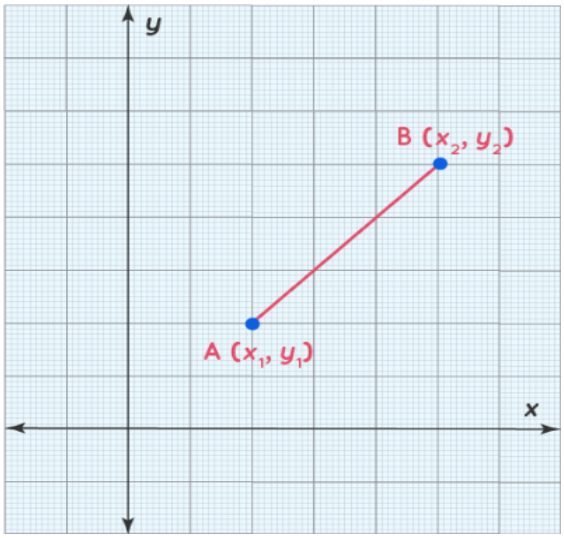

Figure 2: Euclidian Distance

K Nearest Classifier (Input variables for the cluster):

a) Assign K to the number of clusters.

b) K instances are chosen to act as the hubs of the clusters.

c) Considering each input data point:

d) Do the Euclidian distance calculation.

e) The cluster close to the data point should be assigned.

f) Recalculate the centroids and redistribute the clusters' variables.

When there are more than two possible classifications for the target variable, this algorithm is utilized. The only variable that matters in our dataset is age of victim, which is divided into three categories: male, female, and total.

4. RESULT AND ANALYSIS

A. Dataset Description

Survey for Indian Child Crime cases [10]:

DOI : https://doi.org/10.56452/7-5-199

Before moving to main KNN algorithm based on Kaggle dataset from state California we also analysed India Child rates from the year 2002 to 2012. We can see how values increased drastically from 2002 to 2012 year.

Purpose of study crime rates are increasing not decreased or stable at some data point that can be red flag in crime terminology. Hence, to check similar result from diverse data we proposed system with California Male-Female crime rates prediction.

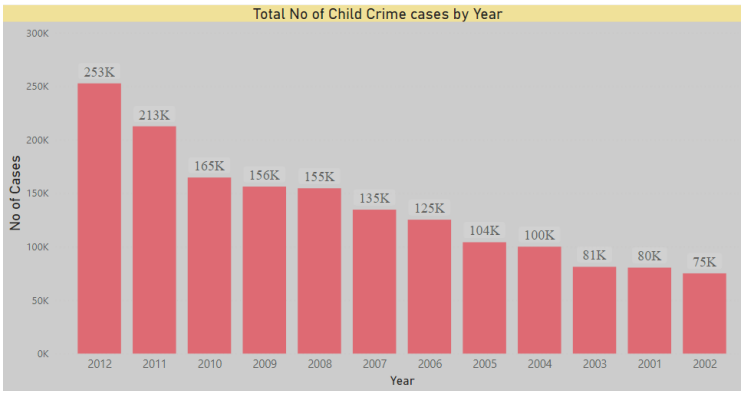

Figure: Total No of Child Crime cases by Year 2002 to 2012 (Ref. Dataset from Data.world)

B. Result

1) In first phase of analysis we used dataset to predict crimes against female by age ratio top 10 tracts in the California state as follow:

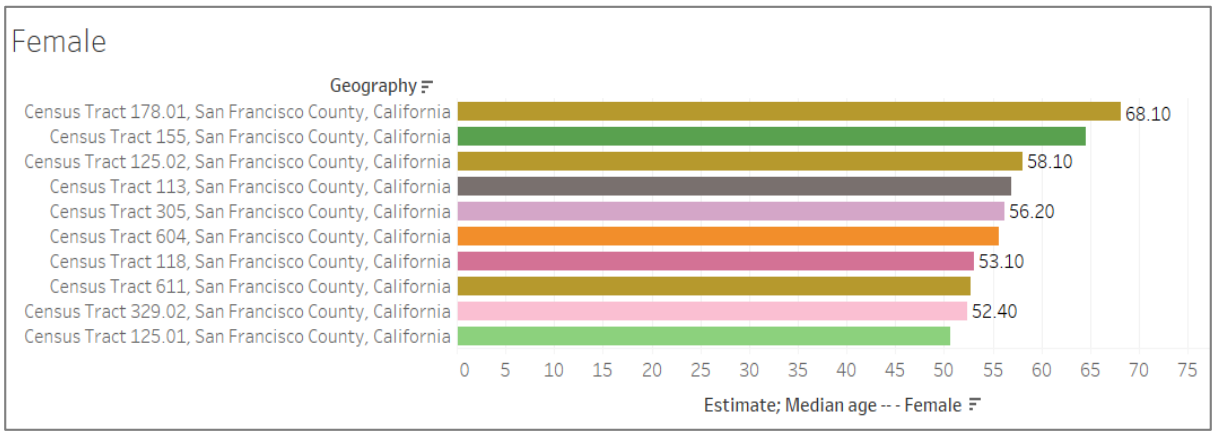

Figure 3. Female Crime prediction by Age

2) Here in second phase, we analysed Male crimes by age top 10 tracts in the California state area as follow:

DOI : https://doi.org/10.56452/7-5-199

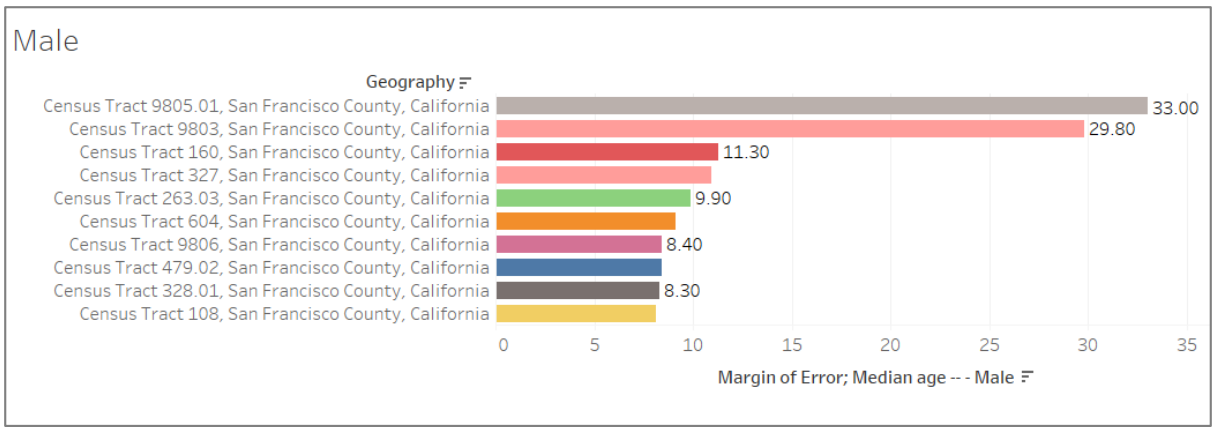

Figure 4 Male Crime prediction by Age

3) In final stage we analysed both Female and Male crime rates as a Total by age index for top 10 tracts in the California state as below:

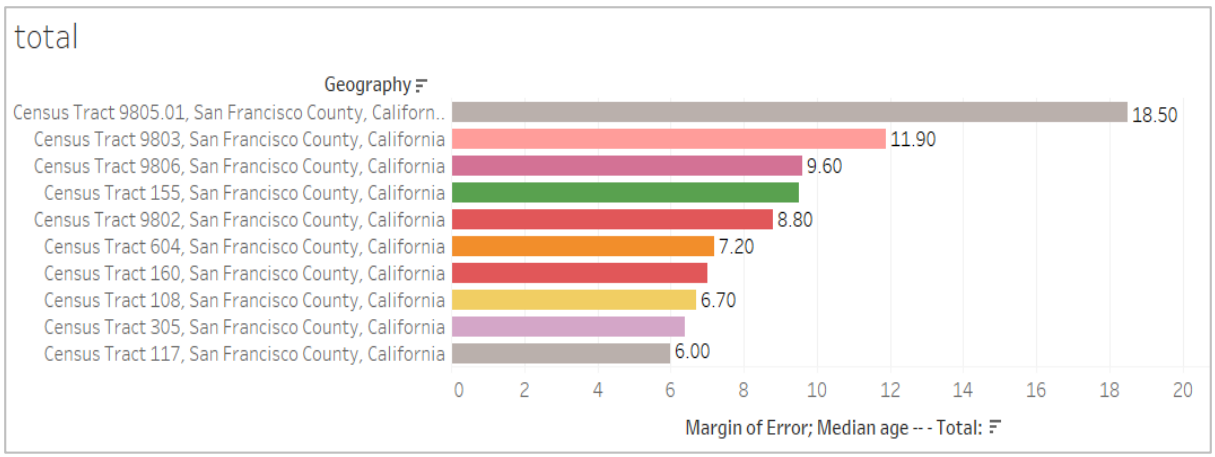

Figure 5 Total Crime prediction by Age

5. CONCLUSION

We develop a crime prediction system using various techniques as machine learning and deep learning technologies, specifically employing the K-Nearest Neighbour classifier and support vector machine. The models analysed show promising accuracy and effectiveness, aiding law enforcement in resource allocation and decision-making. By analysing crime datasets, visualizing data through graphs and charts, and comparing various algorithms to optimize accuracy, the project demonstrates the potential for machine learning applications in the field of crime prevention. Importantly, the system is adaptable to various regions and countries, provided that relevant datasets are available, highlighting its potential for widespread implementation and significant impact on global crime reduction efforts. However, addressing ethical, privacy, and fairness concerns is crucial to avoid biases and discrimination. Continuous collaboration and future research will refine these models and contribute to a safer society.

REFERENCES

[1] Tamir, Azwad, Eric Watson, Brandon Willett, Qutaiba Hasan, and Jiann-Shiun Yuan. "Crime Prediction and Forecasting using Machine Learning Algorithms." International Journal of Computer Science and Information Technologies 12, no. 2 (2021): 26-33.

[2] Shah, Neil, Nandish Bhagat, and Manan Shah. "Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention." Visual Computing for Industry, Biomedicine, and Art 4 (2021): 1-14.

[3] Walczak, Steven. "Predicting crime and other uses of neural networks in police decision making." Frontiers in Psychology 12 (2021): 587943

[4] International Journal of Advanced Research in Computer and Communication Engineering Vol. 10, Issue 5, May 2021 DOI 10.17148/IJARCCE.2021.10537 Copyright to IJARCCE IJARCCE 197 This work is licensed under a Creative Commons Attribution 4.0 International License Crime Prediction Using Machine Learning Approach Prof. Shivaprasad More, Sakshi Mench, Saloni Kuge , Hafsa Bagwan

[5] International Research Journal of Engineering and Technology (IRJET) Crime Prediction and Analysis Using Machine Learning Alkesh Bharati, Dr Sarvanaguru RA.K

[6] Rani A, Rajasree S (2014) Crime trend analysis and prediction using mahanolobis distance and dynamic time warping technique. Int J Comput Sci Inf Technol 5(3):4131–4135

[7] Kim S, Joshi P, Kalsi PS, Taheri P (2018) Crime analysis through machine learning. In: Paper presented at the IEEE 9th annual information technology electronics and mobile communication conference. IEEE, Vancouver 1-3 November 2018.

[8] Duan L, Hu T, Cheng E, Zhu JF, Gao C (2017) Deep convolutional neural networks for spatiotemporal crime prediction. In: Paper presented at the 16th international conference information and knowledge engineering. CSREA Press, Las Vegas 17-20 July 2017.

[9] Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.

[10]https://www.kaggle.com/datasets/jmortin/crime-prediction-sf-dataset?resource=download