International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

PREDICTIVE ANALYTICS FOR CRIME TREND ANALYSIS USING MACHINE LEARNING

[#1]**MEDISHETTI RAMYA SREE,** ***Assistant Professor,***
[#2]**REVELLI KISHOR KUMAR,** ***Assistant Professor,***
Department of Computer Science & Engineering,
Mother Theressa College of Engineering & Technology, Peddapalli, Telangana.

ABSTRACT: Crime is one of the major challenges faced by modern society, making its analysis and prevention a critical priority. This study proposes a systematic approach to crime analysis using machine learning techniques to identify patterns and trends that can support effective investigation. The primary objective is to improve crime-solving efficiency by predicting offender characteristics based on available case information.The research focuses on two key prediction aspects: the gender and age of offenders. For unresolved cases, multiple factors such as year, month, weapon type, and crime details are analyzed. The dataset used for this study was obtained from Kaggle and includes information about victims, offenders, and their relationships.Machine learning models, including Neural Networks, K-Nearest Neighbors (KNN), and Multiple Linear Regression, were applied for model development and evaluation. The results demonstrate the potential of data-driven approaches to assist law enforcement agencies in understanding crime patterns and supporting investigative decision-making.

Keywords: *Crime Prediction, KNN, Decision Tree. Multilinear Regression; K-Neighbors Classifier, Artificial Neural Networks.*

1. INTRODUCTION

The commission of the act constitutes the violation. Someone has betrayed your trust. The law forbids this specific conduct. It is extremely difficult for law enforcement to detect and evaluate criminal conduct that is taking place beneath the surface. Furthermore, a plethora of event-related data is at your fingertips. Consequently, the inquiry should gain from a number of methods. Therefore, the proposed action should help get the criminal situation resolved. Applying machine learning techniques can improve crime analysis and prediction. Regression methods are available inside the machine learning approach. To accomplish the reserch's primary goal, classification procedures are employed. Data analysis frequently makes use of regression methods, particularly multilinear regression, as a statistical tool. Any two numbers can have their relationship evaluated easily using this procedure. With this method, we may use the values of the independent variable to make predictions about the dependent variable. There are various ways to construct classifiers; one of them is the K-Nearest Neighbor classifier. Multiclass target variables are sorted using classifiers. The application of neural

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

networks considerably enhances accuracy. Neural networks rely heavily on their input and output layers. Using these algorithms, we may make educated guesses about the perpetrator's gender, age, and connection to the crime. Hence, the purpose of the model is to aid law enforcement in their investigation. Because of this, it is useful for solving murder mysteries. For comparable studies in the future, this one can serve as a template. An electronic version is available for download on the conference website. Any inquiries you may have regarding the paper's requirements can be sent to the conference publications committee via the contact information provided on the conference website. Please refer to the conference website for details on the submission of your final work.

2.SYSTEM DESIGN

Filtering and wrapping are two machine learning pre-processing techniques that are employed to exclude extraneous information from the acquired data. The data is also made more understandable by reducing its dimensionality. The data is subsequently separated in a subsequent stage. The data used for training and the data used for testing are two separate sets of information. The training and testing datasets both contribute to the learning process.Stage three involves the mapping process. Various types of incidents, as well as days, hours, months, and locations, are denoted by numbers throughout the mapping process. Using the Naive Bayes method, we first check if the traits' relationship is independent. To sort the collected independent features, the Bernoulli Naive Bayes algorithm is employed. The classification of criminological factors is necessary for the analysis of criminal activity within a particular spatial and temporal context. In order to identify the most prevalent crimes, we have integrated geographical and temporal data. To determine how well a prediction model works, one might look at its accuracy rate. To build the prediction model, we utilized Python and Colab, an IDE tailored to data analytics and machine learning applications.

Module Description

A. Data Pre-Processing

To reduce unnecessary violations, pre-processing open source content is crucial. To be included in the dataset, Denver underwent a comprehensive six-year process of collecting crime statistics. The expected outcome is the discovery of the missing integral through the use of machine learning techniques, specifically the filter and wrapper approaches, to the available attribute values. When training and initializing a prediction model, clean data is essential. Data cleansing refers to the process of removing unnecessary information from a set of documents.The relevance of specific features can be evaluated using filtering algorithms. It is essential to think about how a feature relates to the values you are trying to predict. The wrapper method uses a trained prediction model to determine the feature subset's utility. The data is split

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

into a training set and a test set after cleaning.

B,.Mapping

The first thing to do is to separate the details of the incident, such as the crime type and the exact time and date it was committed. In order to make tagging easier, the data is transformed into an integer format. Charts are made from the data that has been analyzed after collection. Because of its strengths in machine learning, Python is chosen to execute the proposed task. Matplotlib makes it easy to make a graph showing the distribution and frequency of criminal conduct. The most prevalent criminal acts can be better anticipated with the help of this graph.

C.Naive Bayes

For crime prediction, Naive Bayes is a popular choice due to its ability to incorporate spatial and temporal data. Because there are a lot of variables that affect the values of the selected criminal traits, the investigation begins by focusing on their independence. Crimes including gang rape, burglary, sexual assault, robbery, murder, chain snatching, armed robberies, cross-state robberies, and robberies serve as training data before being incorporated into the model construction process. Several methods have contested Naive Bayes in the academic literature.

It is common practice to employ the Naive Bayes classifier when trying to ascertain whether attributes of a Gaussian-shaped sample possess continuous values. It is the training data that establishes the normal distribution's mean and standard deviation. Many classifiers deal with categorical data using Multinomial Naive Bayes. In order to operationalize the feature impacts of the provided parameters for crime prediction, the Bernoulli Naive Bayes method is employed..

D. Crime Prediction

The expected crime category is determined by extrapolating the underlying criminal ingredients. After that, the characteristics have an effect on the nominal values.

3. LITERATURE SURVEY

Many illegal acts take place in various settings. Several researchers have laid the groundwork for future investigations into the relationships between crime and monetary indicators including wages, unemployment rates, and levels of education. The K-nearest neighbor (KNN) algorithm and the decision tree methodology are two machine learning models created by Suhong Kim and Param Joshi. Predicting future crime trends and accurately classifying different types of crimes both have an accuracy range of 39% to 44%. The name of this person is Franklin Fredrick.

Data mining is a process that David H. used to improve the dissemination of information by searching through massive datasets. In order to validate newly found patterns, it is essential to compare them to data that was obtained earlier. Shraddha S. Kavathekar employed association rule mining to forecast criminal behavior. Two approaches to machine learning have been recognized: deep neural networks (DNNs)

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

and artificial neural networks (ANNs). A feature-rich dataset enhances a deep neural network's accuracy. With the help of fully linked convolutional layers and deep neural networks (DNN), the prediction model for multi-labeled input identification was developed.

Specifically for Deep Learning methods that include dropout layers, the system was built utilizing the Tenserflow API. According to the results, pre-processing is crucial in situations with a lot of missing data since criminal activity is concentrated in some places instead of being diffused. When it comes to solving problems and making predictions, Artificial Neural Networks (ANNs) mostly depend on trend analysis. Each and every processor in the system works together to create the model. For feature extraction in cloud computing-based data processing, Chandy and Abraham suggested a random forest classifier. Users' IDs, request numbers, arrival times, expiration times, and memory requirements are all viewable pieces of data. Workload prediction is performed utilizing a dataset that was learned and collected during the learning phase, following feature extraction. The system can potentially adapt to user input as it learns the details of the created characteristics.

According to Rohit Patil, MuzamilKachi, PranaliGavali, and Komal Pimparia, the Apriori method can be used to identify structures that occur frequently. The examination also considers the outcomes of the K-means algorithm. The present method is failing because of the increase in criminal activity over the past few years, which makes processing data manually more laborious. This has led to the adoption of state-of-the-art machine learning techniques, such as K-means clustering. In order to establish the best time to intervene, this project will conduct a systematic literature review (SLR) to find out how people find crime hotspots and how effective such strategies are. A reserch of previous studies on the topic of forecasting crime hotspots over time and geography led to the recommendation of using a Systematic Reserchreserch technique. A model for predicting gas transmission pipeline failures was developed by Nasiri, Zakikhani, Kimiya, and Zayed.

The primary goal of the model was to detect rusting. Most prediction models use data from small historical samples or experimental samples. Consequently, we can disregard corrosion that is produced by different environmental variables. Research into various data mining techniques by Nikhli Dubey and Setu K. Chaturvedi made effective identification of likely future crimes conceivable. The employment of a computer procedure grounded in machine learning techniques allows for the categorization of cybercrimes. Because it is easy to install on computers, this method is useful for investigating the incidence of cybercrime in a country. To forecast criminal behavior utilizing feature level data and an appropriate number of parameters, Kang & Kang (year) offered a deep neural network fusion approach.

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

4.SYSTEM ANALYSIS

One possible approach to provide a comprehensive explanation is by utilizing a specific tuple as an illustrative example.

1. Taking into account a tuple
2. {Gateway town, 20th October 2020 , 2: 30 PM, Friday} => {Larceny – a crime involves the theft of a particular's property}

Based on the evidence that has been gathered, it is probable that the following event will take place.

1. {Gateway town} => {Theft has occurred}
2. {October} => {Theft has occurred}
3. {2020} => {Theft has occurred}
4. {2:30 PM} => {Theft has occurred} 5. {Friday} => {Theft has occurred}

After the establishment of the independent event, the conditional probability is calculated. By employing this methodology, it is possible to make anticipations regarding the category of criminal activity. The utilization of symbols

1. m represents Month
2. t represents Time
3. a represents Area
4. d represents Day
5. y presents Year
6. c represents Type

The Formula using the chain in order to find the conditional probability:-

$$P(c|m, y, a, t, d) = [P(m|c, y, a, t, d) * P(y|c, a, t, d) * P(t|d, c) * P(d|c) * P(c)] / [P(m|y, a, t, d) * P(y|a, t, d) * P(a|t, d) * P(t|d)]$$

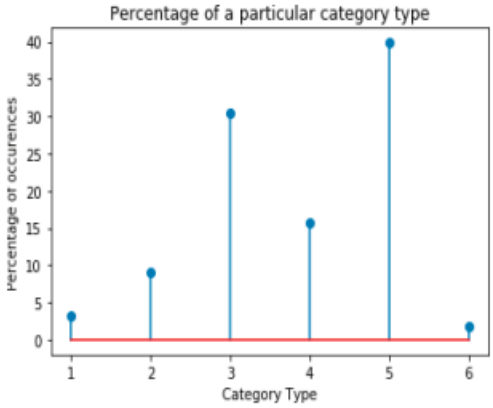

Fig 1. Plotting the highest crime type

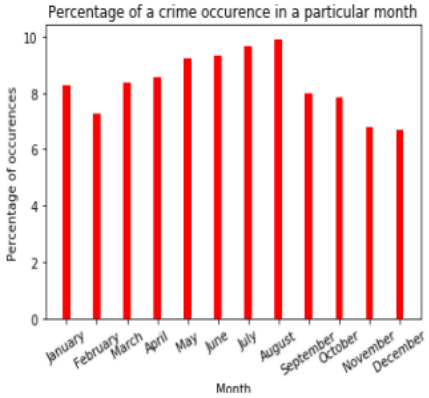

Fig 2. Plotting the highest occurrence month

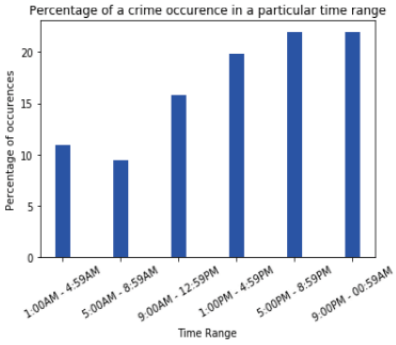

Fig 3. Plotting the highest occurrence time range

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

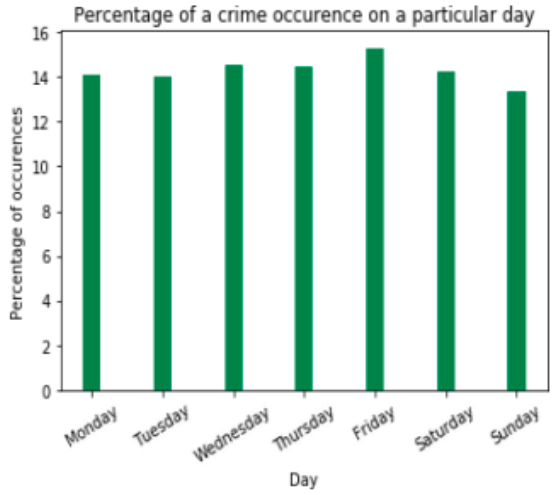

Fig 4. Plotting the highest occurrence day

RESULTS

The goal of the performance evaluation is to improve the forecast's accuracy compared to the prior model used. As it allows the data to be trained on multiple training sets, cross validation is a common technique used during the training phase. The purpose of this research is to determine the efficacy of cross-validation in establishing the reliability of full splits. Python requires data parameters like the name of the model, the target set, and the settings for cross-validation (CV) that help to identify data splits so that you may acquire the right accuracy score. The following step is to determine the average precision's standard deviation and mean. Remarkably, the accuracy has improved by 93.07% compared to earlier prediction models..

EVALUATION METRICS	CROSS VALIDATION
Accuracy	93.07%
Precision	92.53%
Recall	85.76%
F1 score	92.12%

Table1. Performance measure for Naïve Bayes classifier

5. CONCLUSION

The challenge of handling data with both nominal and real values is addressed in this reserch by employing multinomial and Gaussian Naive Bayes (NB) classifiers. Minimal training time is required to generate real-time projections. It also successfully sidesteps the problem of dealing with a continuous collection of target variables, unlike the prior approach. So, Naive Bayesian Classification is a great tool for seeing and anticipating typical criminal actions. A standard set of metrics can also be used to assess the algorithm's performance. Average precision, recall, F1 score, and accuracy are four crucial metrics that must be considered when evaluating an algorithm. When machine learning techniques are applied, accuracy is greatly enhanced.

REFERENCES

[1] Suhong Kim, Param Joshi, Parminder Singh Kalsi,Pooya Taheri, “Crime Analysis Through Machine Learning”, IEEE Transactions on November 2018.

[2] Benjamin Fredrick David. H and A. Suruliandi,“Survey on Crime Analysis and

International Journal of Technovation and Business Insights
Volume 02, Issue 01, January-March 2026
ISSN: 3107-7900, P.CODE: 26/2-1/26
DOI: https://doi.org/10.5281/zenodo.19143475

Prediction using Data mining techniques", ICTACT Journal on Soft Computing on April 2012.

[3] Shruti S.Gosavi and Shraddha S. Kavathekar,"A Survey on Crime Occurrence Detection and prediction Techniques", International Journal of Management, Technology And Engineering , Volume 8, Issue XII, December 2018.

[4] Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no. 02 (2019): 108-118.

[5] Learning Rohit Patil, MuzamilKacchi, PranaliGavali and Komal Pimparia, "Crime Pattern Detection, Analysis & Prediction using Machine", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06, June 2020

[6] Umair Muneer Butt, Sukumar Letchmunan, FadratulHafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, "Spatio-Temporal Peerreviewed journal, published on April 2017.

Crime Hotspot Detection and Prediction: A Systematic Literature Review", IEEE Transactions on September 2020.

[7] Nasiri, Zakikhani, Kimiya and Tarek Zayed, "A failure prediction model for corrosion in gas transmission pipelines", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, (2020).

[8] Nikhil Dubey and Setu K. Chaturvedi, "A Survey Paper on Crime Prediction Technique Using Data Mining", Corpus ID: 7997627, Published on 2014.

[9] Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.

[10]Hyeon-Woo Kang and Hang-Bong Kang, "Prediction of crime occurrence from multimodal data using deep learning",