IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

PREDICTIVE CRIME ANALYTICS: ANTICIPATING CRIME PATTERNS

Dr.S.Jessica Saritha
Assistant Professor, CSE Department, JNTUA College of Engineering Pulivendula
sjsaritha.cse@jntua.ac.in

Abstract :

The integration of machine learning algorithms into crime analysis is revolutionizing law enforcement strategies by predicting crime types and circumstances with unprecedented accuracy. These algorithms use historical crime data and techniques such as Random Forest to classify crime patterns and predict their likelihood of occurring in various areas. By analyzing data obtained through platforms like Kaggle, researchers can better understand recent crimes and promote actions to improve public safety.

Lawmakers, law enforcement, and city planners can use this information to more effectively allocate resources and develop crime prevention plans. Using this information will not only help you identify hotspots of crime, but also help you implement crime prevention measures to reduce the threat to people.By harnessing the power of machine learning, law enforcement agencies can develop strategies to prevent crimes before they happen, rather than focusing on tracking facts. This fight allows society to better deal with security issues, which ultimately leads to a safer and stronger society.

Keywords:

Machine learning algorithms, crime analysis, law enforcement, predictive modeling, Random Forest, historical crime data, crime patterns, public safety, resource allocation, crime prevention, hotspots, proactive strategies, security issues, societal safety

I. INTRODUCTION

Predictive crime analytics uses machine learning algorithms to revolutionize policing and urban planning by providing insight into crime patterns and trends. Using historical crime data and various factors such as geographic area and time, predictive models can predict different types of crime at a particular place and time. This approach allows stakeholders to better allocate resources, prioritize interventions, and improve monitoring opportunities, ultimately improving public safety and prevention strategies. But the field of crime prevention has faced significant challenges, including ethical and privacy issues around the collection and use of data.

Addressing these issues is critical to building community trustandensuring forecast model accountability. Additionally, the nature of the crime poses a problem as background information may be outdated or insufficient for the truth. Combining real-time data with evolving learning technology can help meet this challenge, uncover new trends and adjust strategies accordingly.

Despite these challenges, crime prediction offers many opportunities for evidence-based decision-making, resource allocation and coordination among those involved. By identifying high-risk areas and vulnerable groups, policymakers can implement targeted interventions to reduce crime risk and improve community well-being. Additionally, predictive models will help create more effective and efficient urban planning strategies, thus creating a safer and better urban environment. Collaboration between law enforcement agencies, government agencies, and schools is critical to developing good data, models, and sharing practices to increase confidence, accuracy, and efficiency in crime rate standards.

II. LITERATURE SURVEY

Crime prevention and detection has become an important trend in crime and very difficult to solve crime. Several studies have found different methods for solving crimes that are used in many applications. Such investigations can speed up criminal investigations and help computer systems automatically identify criminals. In addition, rapidly developing technologies can help solve such problems. However, crime patterns are constantly changing and growing.[1].

In contemporary landscape, the surge in criminal activities poses a significant threat to societal well-being and stability. Addressing this challenge necessitates a thorough understanding of crime patterns to proactively respond to such incidents. This study focuses on analysing crime patterns using data sourced from Kaggle, an open platform, to predict recent criminal activities. Emphasizing the identification of predominant crime types, alongside their temporal and spatial occurrences, forms the core objective of this research endeavour. Leveraging machine learning algorithms like Naïve Bayes enables efficient classification of diverse crime patterns, yielding notably high accuracy rates compared to existing methodologies. Through such predictive analytics, this study aims to equip stakeholders with actionable insights for preamptive crime mitigation strategies[2].

Crime prediction is critical for shaping policing strategies and implementing effective crime prevention measures. While machine learning stands as the prevailing predictive method, few studies have systematically compared different machine learning algorithms for this purpose. This paper uses historical data on public property crime from 2015 to 2018 in a large coastal city in Southeast China to evaluate the predictive performance of several machine learning algorithms. The results indicate that the LSTM model outperforms KNN, random forest, support vector machine, naive Bayes and convolutional neural networks when only historical crime data is considered. In addition, the integration of built environment data such as points of interest (POI) and urban road network density as covariates into the LSTM model improves the prediction accuracy compared to the original model based only on historical crime data. Thus, future crime prediction efforts should use both historical data and variables related to criminological theories, recognizing that not all machine learning algorithms are equally effective at predicting crime[3].

III. PROPOSED SYSTEM

In contemporary society, the rise in crime poses significant challenges, disrupting the fabric of communities and nations alike. Understanding crime patterns is essential for

effective analysis and response to criminal activities. Researchers have conducted numerous studies, particularly focusing on Identifying crime hotspots, which aid in classifying and resolving crimes more efficiently.

In recent years, the integration of machine learning algorithms into crime analytics has garnered considerable attention for their ability to forecast crime types and occurrences. This marks the emergence of predictive crime analytics, employing advanced techniques to anticipate and comprehend crime patterns. The focus lies in developing models capable of predicting specific crime types and their likelihood of happening in different locations.

These algorithms leverage historical crime data and pertinent features to train and optimize the models. This paper presents a case of crime pattern analysis that utilizes Kaggle, an open source platform, to forecast recent crimes.Employing machine learning algorithms such as Random Forest enables classification among various crime patterns.

Through rigorous experimentation and validation, this study aims to shed light on the efficacy of machine learning in crime prediction, thereby contributing to the advancement of proactive law enforcement strategies. The insights gained from this research could guide policymakers, law enforcement agencies, and urban planners in devising targeted interventions and resource allocation strategies to bolster public safety and security.

IV. METHODOLOGY

We use the random forest algorithm, a well-known ensemble learning process that is widely recognized for its robustness and adaptability across a variety of learning tasks. Random forest utilizes multiple decision trees, each trained on a sample of data provided by the sampling method. It also varies by considering only some of the traits for each part of the tree, thus increasing the diversityof trees in the group. This difference is important for the performance of the algorithm as it reduces overfitting and increases overall performance.

Parameter tuning is a crucial aspect of optimizing random forest models, particularly in relation to hyperparameters like forest trees., maximum tree depth, and minimum standards required for separation. Implementation is easy using the scikit-learn library in Python, which provides documentation and support as well as efficient use of machine learning algorithms. Using this method, we leverage the power of random forests to achieve our research goals, demonstrate their effectiveness in processing complex data, and provide good predictions.

Entropy:

An entropy is the measure of uncertaintyor confusion in a data set. In machine learning, it evaluates mixed or random patterns or clusters in a dataset. High entropy indicates conflict or inconsistency, while low entropy indicates greater information and predictability.

To calculate entropy you need to consider the distribution of the class list. For most binary classification problems, the formula is: -p * log2(p) (1–p* log2-pp; 1–1–p); where "p" corresponds to the class relative to another class in the data set.For multiclass problems the formula is a little more complicated; The sum -p_i * log2(p_i) for each class "i" is also included.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

Confusion Matrix

By comparing the predictions of a classification model to the actualities and assumptions of different groups, the confusion matrix serves as an uncomplicated evaluation tool. Binary classification involves a confusion matrix that is 2x2 and contains four significant values: true positive (TP), false positive, false negative, and false negative. In certain situations, the model correctly predicts the positive class using true positive values, while false negatives make false predictions of the same class when the correct sign is incorrect. However, a true negative indicates that the prediction of 'false negative' classes is correct, while another nonnegative indicates an incorrect model prediction for such ambiguity: if the class contains bad classes and they are good classes then this implies that it correctly predicted each other.Through the analysis of these results, the model's predictive power can be determined by examining key metrics such as precision and accuracy.Using information gleaned from discrepancy matrices, practitioners can identify areas for improvement and adjust classification models to obtain more accurate and reliable predictions across a wide range of applications and brands.

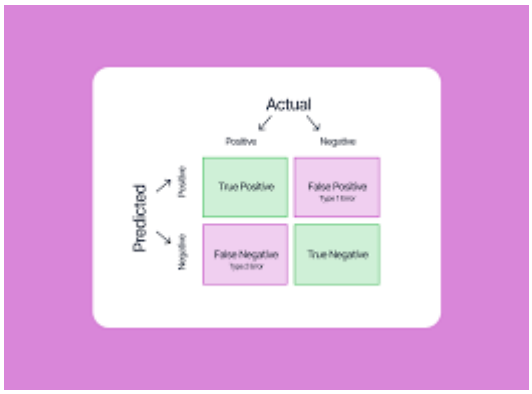

Fig 1: Confusion matrix

Dataset:

Data used in this research has a general methodology that is important for identifying crime patterns and trends. Its distinguishing feature is the latitude and longitude coordinates that provide accurate spatial information about the location of the crime. These geographic coordinates help identify crime hot spots and the distribution of crime in an area. Additionally, including date, time, and year behavior provides temporal information on crime patterns over time, allowing researchers to view seasonal changes in crime, daily patterns, and long-term patterns.

In addition, the database contains a variable that indicates the type of crime associated with each recorded case. These qualities are important in crime analysis and allow researchers to identify and categorize different crimes. By examining the frequency and distribution of specific crimes, such as theft, assault or violence, researchers can better understand crime policies and prioritize interventions accordingly. Overall, the multiple nature of these data (such as spatial, temporal and categorical features) provides a good basis for rigorous crime analysis and predictive modeling, ultimately helping to ensure legal protection and social security measures.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

V. IMPLEMENTATION

Monitoring the system forCrimethrough data mining techniques begins by collecting data from trusted sources such as law enforcement, government records, or open data platforms such as Kaggle to ensure that the dataset has a significant time difference capture. crime patterns and trends. The next stage involves collecting data, cleaning it to eliminate duplicate entries and preserving missing values, as well as transposing changes to numbers using single-bit coding.Additionally, numerical features are normalized or scaled to ensure uniformity of product distribution. After that, a specific selection is made, including statistical data research, to determine the characteristics of the crime that occurred, and the statistical method or information can be recorded. It is used to select specific data for model training. For effective representation, consider external data such as demographics, socioeconomic indicators, and geographic features. Once the features are selected, the next stage will involve model selection and training, where we specifically measure random forest parameters for crime prediction. Use cross-validation methods to test your model's performance and use grid search or random searchto optimize hyperparameters to improve its stability and accuracy. As well as the effectiveness of the training model using temporal or cross-sectional technique to ensure its consistency and reliability in the real situation. Finally, predictive models are identified and visualized to identify key factors contributing to the occurrence of crime, identified crime hotspots, time trends, and predictive models to providerecommendations to stakeholders such as law enforcement, policy makers, and urban planners.

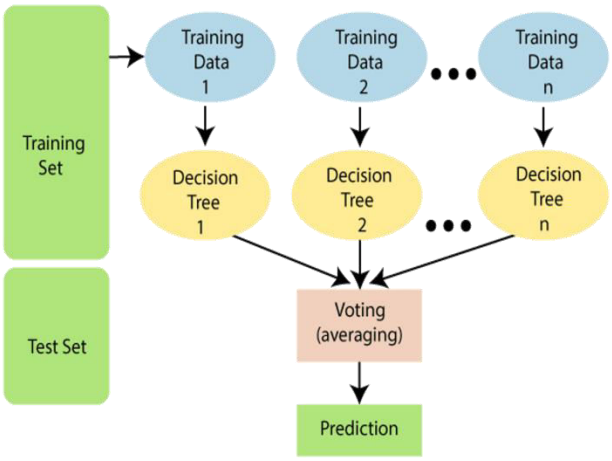

Fig 2: Random forest Algorithm

VI. RESULT AND ANALYSIS

The results and analyses presented in this article highlight the importance of a data-driven approach to crime analysis and demonstrate the potential of machine learning, especially random forests, to improve law enforcement strategies and public safety plans.

The user interface for crime vulnerability prediction allows users to input various parameters including latitude, longitude, date, day, time, and year. These inputs serve as key factors for the system to analyze and predict the vulnerability of a specified location to different types of crime at the provided time. Upon user input, the system initiates a data processing phase where it retrieves relevant information from the dataset based on the entered

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

parameters. This dataset, compiled from reliable sources like law enforcement agencies or open data platforms, provides past crime data along with other relevant elements like location coordinates, timestamp numbers, and types of crimes..

Once the data is processed, the system employs a sophisticated crime prediction algorithm, such as a trained Random Forest model, to analyze the input parameters andpredicts the probability of various crimes in a given place and time. Leveraging machine learning techniques, the algorithm considers historical crime patterns, temporal trends, and spatial correlations to generate accurate predictions. The results of the prediction are then displayed to the user in a comprehensible format.

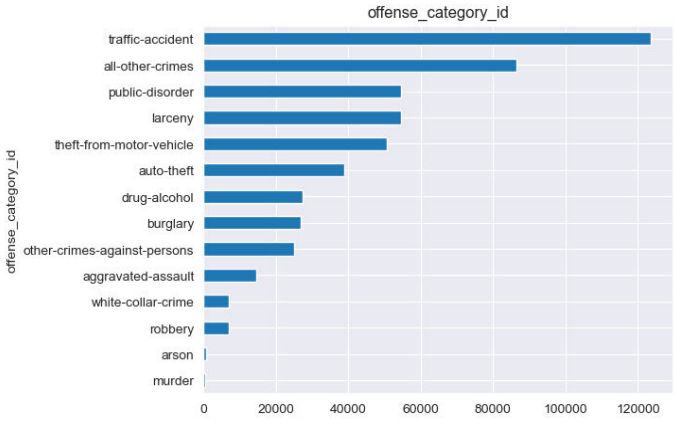

Fig 3: Categories of crimes and

Occurrence

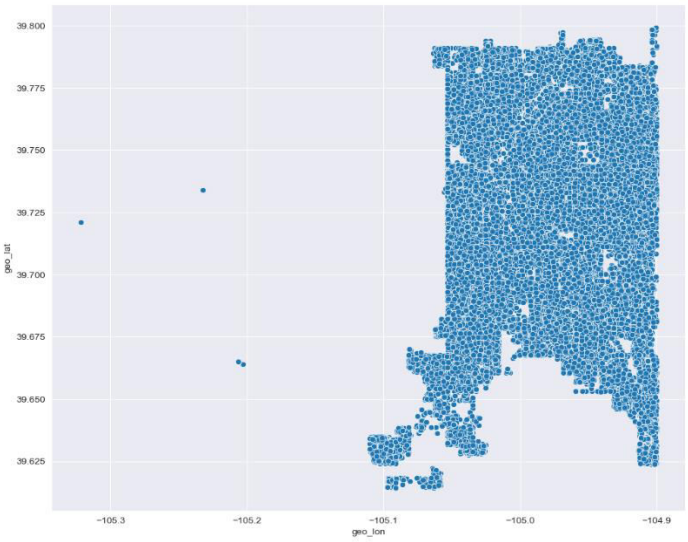

Fig 4: Spatial Distribution of all crimes

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

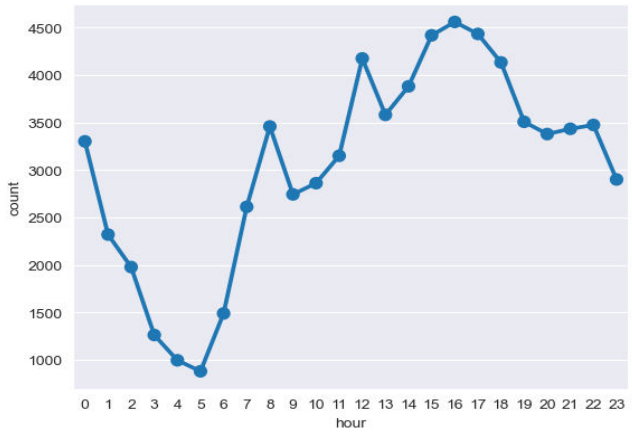

Fig 5: Diurnal occurrence of crimes

Fig 6: Spatial Distribution of various crimes

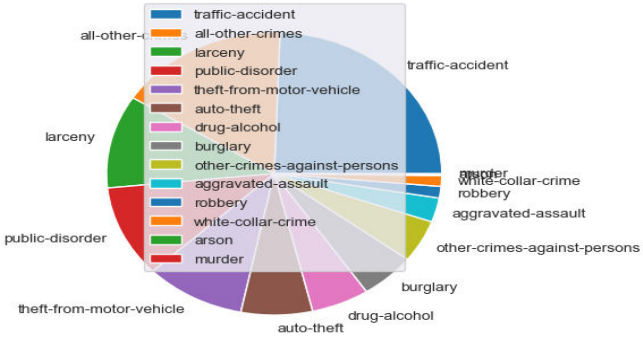

Fig 7: Distribution of various crimes

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 10, Iss 8, 2021

Fig 8 : Home Page

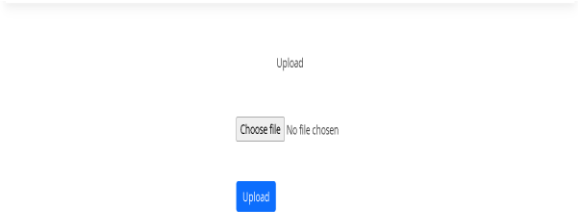

Fig 9 : Upload page to upload dataset

Fig 10 : Predicting page to predict vulnerability of place

VII. CONCLUSION

In summary, our project aims to use machine learning, specifically the random forest algorithm, forecasts the likelihood of different crimes or events occurring in a particular

timeframe and place.We use latitude, longitude, date, time, and year to construct a predictive model of crime using historical data.Through a user-friendly web interface, individuals can enter location and time details to obtain predictions on the likelihood of various types of events in the area. The random forest algorithm was chosen because it is suitable for processing large data sets with dimensions and storing complex interrelated data. By training on historical data, our model learns patterns and relationships betweenvariables, allowing it to predict future events.

VIII. FUTURE SCOPE

Many future improvements could improve the performance of our prediction mod els. First, continuous search and optimization of features can improve predictions. This may i nvolve integrating other data such as weather, socioeconomic indicators or urban context. Ad ditionally, testing different machine learning algorithms and techniques can optimize the mod el's performance. Integrated systems, deep learning or more selective methods may provide i mprovement opportunities.

The current forecast is also worth investigating. Creating a system for instant data integration and updates can enable a more responsive and effective counterterrorism response . It is also important to regularly evaluate and validate predictive models. Regularly evaluate t he performance and validity of models across different datasets and geographies to ensure the y are reliable and generalizable.

REFERENCES

[1] U. Thongsatapornwatana, ``A survey of data mining techniques for analyzing crime patterns," Jan. 2016, pp. 123128.
[2] N V Keerthana, G Ranjitha, S Yuvarani,N Kanimozhi,, G S Pavithra "CRIME Type and Occurrence Prediction Using Machine Learning Algorithm"

[3] LIN LIU, JIAKAI JI,XU ZHANG, AND LUZI XIAO "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots".

[4] E. Poulsen andM. Felson, ``Simple indicators of crime by time of day," Int. J. Forecasting, vol. 19, no. 4, pp. 595601, Oct. 2003.
[5] J.-P. Nadal andH. Berestycki, ``Self-organised critical hot spots of criminal activity," Eur. J. Appl. Math., vol. 21, nos. 45, pp. 371399, Oct. 2010.
[6]Zaheer Abbass "A Framework to Predict Social Crime through Twitter Tweets By Using Machine Learning" 03-05 February 2020
[7] Y. B. Chen ,W. H. Li, and L.Wen, ``Application of improved GA-BP neural network model in property crime prediction," Geomatics Inf. Sci. Wuhan Univ., vol. 42, no. 8, pp. 11101116, 2017.
[8]Christoffer gahlin "Crime hotspots: An evaluation of the KDE spatial mapping techniques".
[9] C.P. JOHNSON, ``Crime Mapping and Analysis Using GIS,"

[10] W.W. Liao, J. Liu, andL. Lin,W, ``Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction," Prog. Geogr., vol. 37, no. 6, pp. 761771, 2018.
[11]Fei Yi∗, Zhiwen Yu, ``An Integrated Model for Crime Prediction Using Temporal and Spatial Factors,"
[12]Hong Chen “Improved naive Bayes classification algorithm for traffic risk management”, 2021
[13] “Enhancing Urban Safety and Security. Global Report on Human Settlements” UN-Habitat, Nairobi, Kenya, 2007.
[14] A. Tinghua and Y. Wenhao, ``The visualization and analysis of POI features under network space supported by kernel density estimation," Acta Geodaetica et Cartographica Sinica, vol. 44, no. 1, pp. 8290, 2015.
[15] L. Xiao, K. Liu,G. Song, D. Long, S. Zhou, and K. LiuS. Zhou, ``Impact of residents' routine activities on the spatial-temporal pattern of theft from person," Acta Geography Sinica, vol. 72, no. 2, pp. 356367, 2017.
[16]S. H. Zhou, L. Liu, and C. Xu, ``The comparison of predictive accuracy of crime hotspot density maps with the consideration of the near similarity: A case study of robberies at DP Peninsula," Scientia Geographica Sinica, vol. 36, no. 1, pp. 5562,2016.-
[17] S. D. Johnson andA. Sagovsky, ``When does victimisation occur?" Australia. New Zealand J. Criminol., vol. 40, no. 1, pp. 216, 2007.
[18] E. Poulsen andM. Felson, ``Simple indicators of crime by time of day," Int. J. Forecasting, vol. 19, no. 4, pp. 595601, Oct. 2003.