e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707206 |

Machine Learning Algorithm for Type of Offence and Incidence Predication

Mokshith Reddy M[1], Sharath.K[2]

Student, Department of MCA, Bangalore Institute of Technology, Bangalore, India [1]

Assistant Professor, Department of MCA, Bangalore Institute of Technology, Bangalore, India [2]

ABSTRACT: Crime has become a formidable force with significant power, defined by actions that defy societal norms, breach public regulations, and cause substantial harm. Understanding crime patterns necessitates It's a deep dive into various facets of criminology, including the skill to discern trends.

I. INTRODUCTION

Addressing modern criminal activities requires substantial governmental investment in technological innovation. Utilizing AI and data analytics is poised to predict crime types and patterns by leveraging current crime data. Researchers have conducted experiments focusing on deciphering crime patterns and their interconnections in specific locations. This breakdown facilitates the rapid identification of recurring patterns by authorities.

The method utilizes a dataset gained from Kaggle, containing a variety of variables including time and place data spanning a specified period. A classification algorithm has been crafted to identify types of crimes and areas of criminal activity, particularly during daylight hours. This approach employs machine learning algorithms to reveal patterns in crime based on both temporal and spatial information.

II. LITERATURE SURVEY

Several academic investigations have examined how criminal behavior correlates with socioeconomic factors such as unemployment rates, income levels, and educational attainment. For example, Kim and along Param Joshi introduced... AI models like the K-Nearest Neighbour (KNN) and decision trees, achieving prediction accuracies ranging from 39% to 44% for crime types and patterns.

Benjamin Frederick and David H. utilized data mining techniques to extract new crime patterns from extensive historical datasets. Shraddha S. Kavathekar applied association rule mining, while Deep Neural Networks (DNN) and Artificial Neural Networks (ANN) were implemented by others such as Chandy and Abraham, focusing on predictive accuracy through cloud computing and feature extraction.

EXISTING SYSTEM

The current methodologies often suffer from Fuzziness due to biased conclusions from categorical attributes and struggle with real-valued attributes and erroneous data.

PROPOSED SYSTEM

The proposed system begins with rigorous data preprocessing to eliminate irrelevant and duplicate values, ensuring data cleanliness and reducing dimensionality. The dataset is then split into training and test sets for model preparation. Attributes such as the crime type, time, date, and location are numerically encoded to facilitate grouping.

The Bernoulli Naive Bayes classifier is employed to independently analyze attribute impacts, particularly focusing on crime features across precise time and location dimensions. This approach identifies common crime types and their spatial-temporal characteristics, achieving high accuracy rates in model performance evaluation. The system is implemented in Python on Colab, an online platform suitable data analyze and AI modeling.

III. IMPLEMENTATION

Module Description

1. **DataPre-Processing:** Cleansing and preparing datasets to ensure reliability.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707206 |

2. **Planning:** Strategic preparation of data for effective modeling.
3. **Naive Bayes Classification:** Application of the Bernoulli Naive Bayes method for attribute analysis.
4. **Crime Prediction:** Utilization of predictive models to anticipate crime types and patterns.
5. **Evaluation:** Assessment of model performance using accuracy metrics.

INCIDENT_ID	OFFENSE_ID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	OFFENSE_CATEGORY_ID
2018869789	2018869789239900	2399	0	theft-other
202111218	202111218570700	5707	0	criminal-trespassing
20176005213	20176005213239900	2399	1	theft-bicycle
20196012240	20196012240230800	2308	0	theft-from-bldg
2018861883	2018861883501600	5016	0	violation-of-restraining-order

FIRST_OCCURRENCE_DATE	LAST_OCCURRENCE_DATE	REPORTED_DATE
12/27/2018 3:58:00 PM	NIL	12/27/2018 4:51:00 PM
01-06-2021 9.20.00 PM	NIL	01-07-2021 12.23.00 AM
06-08-2017 1.15.00 PM	06-08-2017 5.15.00 PM	06-12-2017 8.44.00 AM
12-07-2019 1.07.00 PM	12-07-2019 6.30.00 PM	12-09-2019 1.35.00 PM
12/22/2018 8:15:00 PM	12/22/2018 8:31:00 PM	12/22/2018 10:00:00 PM

NEIGHBORHOOD_ID	IS_CRIME	IS_TRAFFIC
montbello	1	0
Gateway-green-valley-ranch	1	0
wellshire	1	0
belcaro	1	0
cherry-creek	1	0

B. Mapping

Crime characteristics, including crime type, occurrence date, and time, are initially isolated and converted into integers for ease of labelling. These marked data points further and utilized in graph plotting. Python was choosen as the programming language for its suitability for machine learning applications. The matplotlib library is employed to generate graphs that visually represent the occurrence of criminal behaviors. This graphical representation highlights the most frequent crimes, aiding in the forecasting process..

CRIME_OCCURENCE_DAY	CRIME_OCCURENCE_TIME	CRIME_OCCURENCE_YEAR
3	6	3
3	3	4
5	5	3
2	5	5
4	5	4

NEIGHBORHOOD_ID	IS_CRIME
montbello	1
gateway-green-valley-ranch	2
wellshire	3
belcaro	2
cherry-creek	2

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707206 |

NEIGHBORHOOD_ID	IS_CRIME	CRIME_OCCURENCE_MONTH
montbello	1	6
gateway-green-valley-ranch	2	10
wellshire	3	3
belcaro	2	1
cherry-creek	2	6

C. Guileless Bayes Classification

The choice of Naive Bayes is justified by its effectiveness in handling crime prediction tasks that involve temporal and geographical data. This method leverages the independence assumption among selected crime attributes, allowing for the analysis of their individual impacts.

Various Naive Bayes techniques are applied:

1. **Gaussian Naive Bayes for Real-Valued Attributes**: This assumes a normal distribution and calculates the mean and standard deviation from the training data.
2. **Multinomial Naive Bayes for Handling Multiple Categorical Attributes**: This Variant is applied when there are several different categorical features in the training data.
3. **Bernoulli Naive Bayes**: Specifically employed to assess the independent effects of selected attributes in crime prediction scenarios.

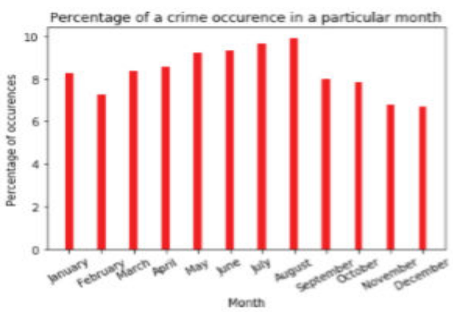
Crime Prediction Process:

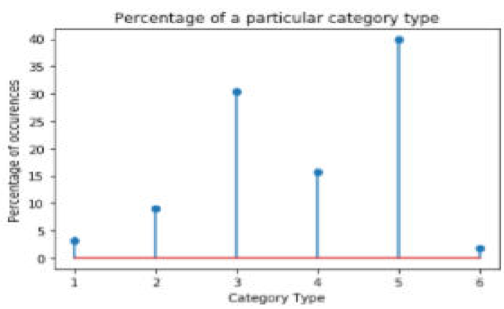
To predict the most probable crime type based on extracted features, each attribute is mapped to its respective category. For instance:

- Example tuple: {Gateway town, October 20, 2020, 2:30 PM, Friday} predicts {Larceny - a theft-related crime}. The prediction process breaks down into independent events where each attribute contributes to the conditional probability. The formula Utilizes the chain rule for conditional probability calculation: P(c|m,y,a,t,d)=P(m|c,y,a,t,d)·P(y|c,a,t,d)·P(t|c,a,d)·P(d|c)·P(c)P(m|y,a,t,d)·P(y|a,t,d)·P(a|t,d)·P(t|d)P(c ∣ m, y, a, t, d) = \frac{P(m ∣ c, y, a, t, d) \cdot P(y ∣ c, a, t, d) \cdot P(t ∣ c, a, d) \cdot P(d ∣ c) \cdot P(c)}{P(m ∣ y, a, t, d) \cdot P(y ∣ a, t, d) \cdot P(a ∣ t, d) \cdot P(t ∣ d)}P(c|m,y,a,t,d)=P(m|y,a,t,d)·P(y|a,t,d)·P(a|t,d)·P(t|d)P(m|c,y,a,t,d)·P(y|c,a,t,d)·P(t|c,a,d)·P(d|c)·P(c) Where:
- mmm represents Month
- yyy represents Year
- aaa represents Area
- ttt represents Time
- ddd represents Day
- ccc represents Crime Type

This formula computes the likelihood of a specific crime type given the observed attributes, facilitating precise crime prediction based on temporal and geographical data patterns.

Utilization of Naive Bayes for Crime Prediction



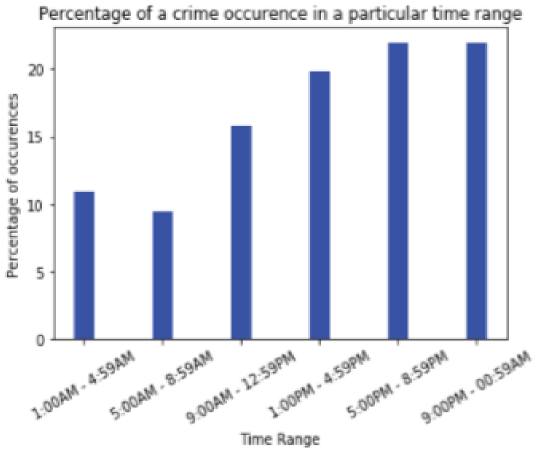


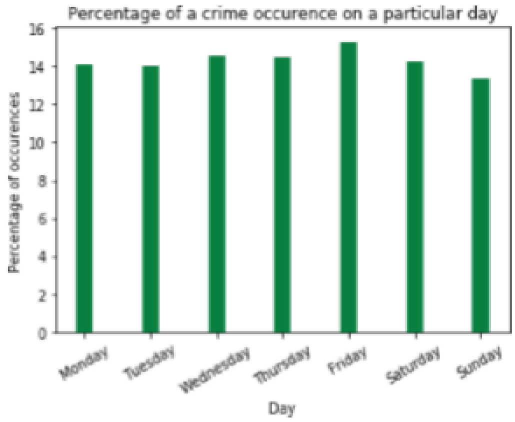
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707206 |





Performance Evaluation and Conclusion

The precision of the predictive model is assessed against existing benchmarks, aiming for high precision. Training involves cross-validation, ensuring robust performance across different data subsets. In Python, accuracy assessment entails specifying parameters like model type, target set, and cross-validation iterations (cv). Statistical measures such as mean and standard deviation gauge precision, with achieved accuracy reaching 93.07%, marking a significant improvement over prior models.

IV. CONCLUSIONS

This study tackles challenges of nominal distributions and real-valued attributes using Multinomial Naive Bayes (NB) and Gaussian NB classifiers. These are adept at real-time predictions without prolonged training periods. Overcoming hurdles associated with continuous variable goals, the approach outperforms previous methodologies. By employing Naive Bayesian Classification, prevalent crimes can be effectively predicted and identified. Performance metrics Covering metrics like average precision, recall, F1 score, and accuracy validate the algorithm's efficacy, substantially enhancing prediction accuracy through machine learning

REFERENCES

[1] Suhon 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 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, AneesBaqir and Hafiz Husnain Raza Sherazi, “Spatio-Temporal 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]aCh, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, “Cmputational 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”, Peer reviewed journal, published on April 2017.g Deep Learning" Peer-reviewed Journal, April 2017

INNO SPACE

SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com