JOURNAL OF DATA SCIENCE | Vol.2024:24
eISSN:2805-5160

Machine Learning Applications in Offense Type and Incidence Prediction

Balaji R[1], Manjula Sanjay Koti[1], Harprith Kaur[2]

[1]Dayananda Sagar Academy of Technology and Management, Bangalore, India
[2]INTI International University, Nilai, Malaysia

Email: balajiragavendracpt@gmail.com, manjula1705vision@gmail.com

Abstract

In today's rapidly evolving world, detrimental behaviour has undeniably emerged as a significant factor leading to the downfall of individuals and communities. The rising prevalence of such behaviour creates substantial disruptions within a country's population, affecting social stability and economic progress. To mitigate the impact of these harmful actions, it is crucial to identify and address them promptly and effectively. This study evaluates specific patterns of detrimental behaviour using data from Kaggle to predict and analyze prevalent negative behaviours. Recent incidents of theft, for example, have underscored the importance of understanding the most common types of misconduct, as well as their timing and locations. We can develop targeted strategies to prevent and respond to such incidents by analyzing these patterns. Artificial Intelligence (AI) techniques encompass various computational methods and algorithms designed to enable machines to perform tasks that typically require human intelligence. These techniques are used in various applications, from natural language processing to image recognition, and offer powerful tools for behavioral analysis. This project employs advanced AI techniques, such as Naive Bayes, to model and identify patterns in detrimental behavior. Naive Bayes, a probabilistic classifier based on Bayes' theorem, is particularly effective in handling large datasets and making accurate predictions. By applying this algorithm, the study achieves a high level of precision in predicting various types of detrimental behavior, enabling a better understanding of their underlying patterns. This knowledge can inform the development of more effective prevention and intervention strategies, ultimately contributing to the reduction of harmful behaviors and the enhancement of community well-being.

Keywords

Machine Learning, Offense Type, Incidence Prediction, Artificial Intelligence

Introduction

Wrongdoing has evolved into a significant strong force that is now thought to have an increased level of power. When a movement disregards the norm, disregards public power rules, and is

Submission: 3 May 2024; **Acceptance:** 25 July 2024

highly unsettling, it is regarded as poor behavior (Kim, S., at. Al, 2018). Difficulty producing plan evaluation necessitates an emphasis on a few topics of criminal science as well as the ability to offer strategies. To supervise a piece of such hoodlum workouts, the government should commit a significant lot of energy and time to collecting growth (David, H., & Suruliandi, A. 2017). As a result, the use of artificial intelligence methods and data is meant to predict the kind of negative conduct and plans (Gosavi, S. S., & Kavathekar, S. S., 2018). It implements the current poor behavior data objectives and evaluates the type of unacceptable conduct and incidence about where and when it occurred. Researchers went through a series of Tests to assist them in unraveling the poor behavior patterns and their connections within an area. A section of the areas of interest has been separated into a clearer method of coordinating the negative behavior plans. Given the suggestions, the professionals will desire to distinguish them more quickly. The technique takes use of an array of data supplied from the freely available Kaggle dataset that describes various factors as well as the reality that they occur over a specific period. We established an order computation that helps in defining the type of heinous behavior and central characteristics of heinous crimes done at a specific time, such as day (Chandy, A. 2019). With the assistance of its association with the provided fleeting and spatial information forces Machine learnings computations to regard as matching lawbreaker designs (Patil, R., et.al., 2020).

Methodology

The methodology applied to this study consists of these steps: Information pre-handling, planning, innocent Bayes's characterization, wrongdoing expectation, and assessment

A. Information Pre-Processing

To avoid unnecessary infringement, information obtained from free sources should be pre-processed (Butt, U. M., et.al., 2020). The data set was chosen for the city of Denver because it includes considerable misconduct information stretching back more than six years. The artificial intelligence approach channels and cover are proven to be the fundamental in resolved property evaluations (Zakikhani, K. et.al, 2021). The setup of an assumption model, as well as the execution of the begun communication, necessitate data cleansing. I finished isolating the situation and removing extraneous values from both datasets. Isolating approaches aid in determining the element determination considering the link with the ward values. The forced covering approach is used to estimate how useful a component subset is by creating a forecast model on it (Dubey, N., & Chaturvedi, S. K. 2014). This can be seen in the table below.

Table 1. Dataset Collection

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
201960122400	20196012240230800	2308	0	theft-from-bldg
2018861883	201886188350160	5016	0	violation-of-restraining-order

B. Mapping

The key details of the wrongdoing are first contained, such as the type of terrible behavior, the time of the misbehavior, and the hour of the event in question. It is then converted entirely to a number to operate with imprinting. The displayed data is therefore broken down and used in graphs. plotting. Python was chosen as the programming language to complete the intended objective since it is suitable for artificial intelligence. The item matplotlib is used to create a chart that depicts the occurrence of criminal behavior. The most frequent breaches can be shown on the diagram, which aids in the process of forecasting.

C. Guileless Bayes Classification

They are utilized for the development of an example by giving preparation based on wrongdoing data linked to burglary, theft, and other crimes. murder, sexual manhandling, equipped burglary, chain grabbing, assault, and roadway burglary. A piece of Nave Bayes' lengthy techniques has been deduced.

1. Genuine esteemed trait selection is linked to Gaussian Nave Bayes. In any event, it is expressed as predicted circulation, this is followed by computing the average variation and mean derived from supplied data (Ch, R.,et.al., 2020).
2. It is ostensibly multi-ostensible. For multiple classifiers that compare to the obvious cut highlights in the prepared worth, Nave Bayes is used.
3. Bernouille Nave Bayees is used to find out the autonomous component effects for a selected ascribes for the wrongdoing expectation (Kang, H. W., & Kang, H. B., 2017).

D. Performance Analysis

By widening the upheld wrongdoing highlights, the typical wrongdoing type is expected. Following that, the ingredients are applied to seeming attributes. using the singles tuple for an occasion, it is possible to understand it clearly.

Considering A Tuple:

1 {Gateway town, twentieth October 2020, 2:30 PM, Friday} => {Larceny - a wrongdoing includes the burglary of a specific's property}

Taking into account plausible events in light of the included extricated:

2. {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}

In addition, the free event has been framed, and the contingent likelihood has been computed. We could predict the misbehaviour type by doing so. The following figures 1,2,3 and 4 illustrate the highest crime type highest occurrences in a month, the time range of highest crime occurrences, and the highest crimes in days.

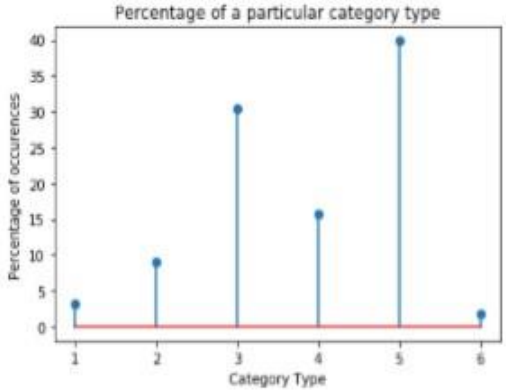

Figure 1. Highest crime type

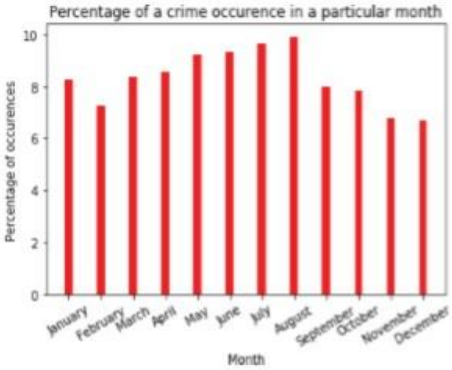

Figure 2. Highest occurrence month

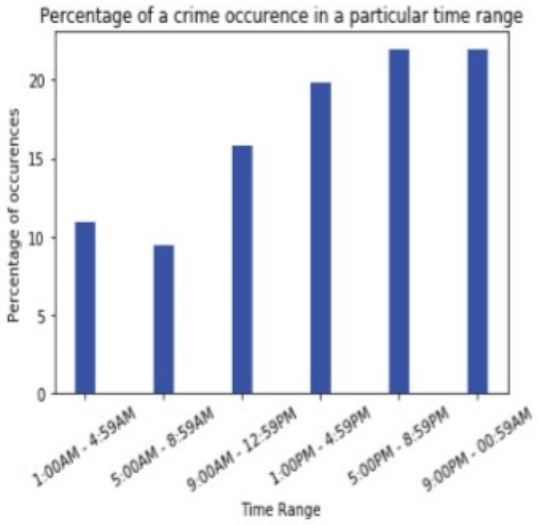

Figure 3. Highest Occurrence Time Range

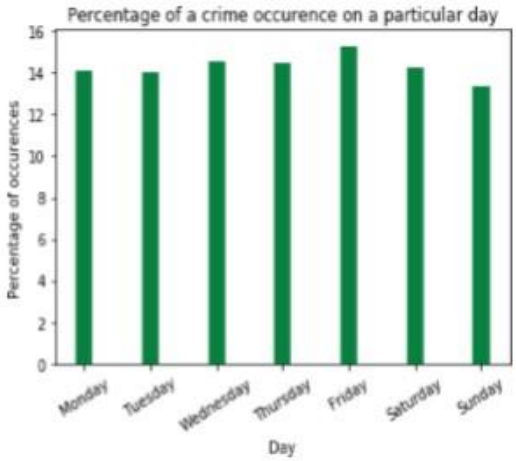

Figure 4. Highest Occurrence Day

F. Evaluation

When contrasted with the previous model, the chosen expectation's presentation obtained an unusually high degree of precision. Cross-validation is used to prepare the information, allowing it to be produced on a variety of preparation sets of data. It allows you to assess the accuracy of most of the components of the cross-approval displayed. To calculate the accuracy number in Python, we must enter data disputes such as model name, target set, and so forth. and cv, which aids in displaying the division event. Finally, the average and standard of normal accuracy are not set in stone. A precision of 93.07 percent was achieved, which is an enormous boost over previous forecast models. Table 2 below shows the cross-validation vales.

Table 2. Cross Validation

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

Conclusion

The challenge of dealing with nominal distributions and with this focus, genuine prestigious credit is obtained by utilizing two classifiers: Multi-ostensible It also addresses the challenge of controlling an ongoing realistic arrangement of elements, something previous endeavors couldn't manage. As a result, the most generally recognized crimes could be predicted or discriminated against using Nave Bayesian Characterization. The calculation's display is also estimated using a few conventional metrics. The calculating evaluation is primarily concerned with metrics, such as normal accuracy, review, F1 score, and precision. The precision of worth may be greatly advanced by combining AI. It has a few cutoff points, regardless of how it approaches a problem for previous work. When there are no class marks, the chance for evaluation is zero. The probable future growth of the suggested usage of more AI group models expands the exactness of wrongdoing expectations and works on the nature of the information. Performance in general It facilitates the setting up of outstanding assistance. After passing the exam, you will want to read up for future advancement. Consider selling data locations in the neighbourhood to see whether there's a link between a country's income levels and the level of crime in the areas closest to it.

References

Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018, November). Crime analysis through machine learning. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 415-420). IEEE. https://doi.org/10.1109/iemcon.2018.8614828

David, H., & Suruliandi, A. (2017). SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES. *ICTACT journal on soft computing*, *7*(3). https://doi.org/10.21917/ijsc.2017.0202

Gosavi, S. S., & Kavathekar, S. S. (2018). A survey on crime occurrence detection and prediction techniques. *International Journal of Management, Technology And Engineering*, *8*(XII).

Chandy, A. (2019). Smart resource usage prediction using cloud computing for massive data processing systems. *J Inf Technol*, *1*(02), 108-118. https://doi.org/10.36548/jitdw.2019.2.006

Patil, R., Kacchi, M., Gavali, P., & Pimparia, K. (2020). Crime Pattern Detection, Analysis & Prediction using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*, *7*(6), 2035-2038.

Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., & Sherazi, H. H. R. (2020). Spatio-temporal crime hotspot detection and prediction: a systematic literature review. *IEEE access*, *8*, 166553-166574. https://doi.org/10.1109/access.2020.3022808

Zakikhani, K., Nasiri, F., & Zayed, T. (2021). A failure prediction model for corrosion in gas transmission pipelines. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *235*(3), 374-390. https://doi.org/10.1177/1748006x20976802

Dubey, N., & Chaturvedi, S. K. (2014). A survey paper on crime prediction technique using data mining. *Int. J. Eng. Res. Appl*, *4*(3), 396-400.

Ch, R., Gadekallu, T. R., Abidi, M. H., & Al-Ahmari, A. (2020). Computational system to classify cybercrime offenses using machine learning. *Sustainability*, *12*(10), 4087. https://doi.org/10.3390/su12104087

Kang, H. W., & Kang, H. B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, *12*(4), e0176244. https://doi.org/10.1371/journal.pone.0176244