AN ENHANCED MODEL FOR CRIME RATE ANALYSIS USING MACHINE LEARNING

G THANUJA [1] B RANI PRASANTHI[2] JAINA SAICHAND [3] P.SRINIVASULA REDDY [4]
[1234] ASST. PROFESSOR, DEPARTMENT OF ARTIFICIAL INTELLIGENCE,
[1,2,3,4] SRI MITTAPALLI COLLEGE OF ENGINEERING

Abstract

Consistently rising crime rates have a negative impact on a nation's economic progress. Reducing crime rates has become an urgent priority, and it is one of the most pressing concerns facing our society. Determining an improved strategy to reduce crime rates requires first identifying the many components and event linkages of criminal activity. In order to do this, it is necessary to maintain a database that details a variety of crimes along with specific information such as location, time, kind of crime, and so on. In order to reduce crime rates in India, this investigation focusses on how AI computations might be designed and analysed. Determining the example relations among enormous arrangements of data has been more easier with the use of AI techniques. Based on the location where it has successfully happened, this exploration mostly depends on making a prognosis on the sort of crime that may occur. The use of AI has allowed for the development of a model via the creation of an informative index that has undergone data cleansing and transformation. With the help of information perception, one may investigate the informational index and its attributes. Every component is being identified and captured. In order to safeguard society, potential dangers are being identified and preventative actions are being prepared. This work has achieved many grouping calculations, improvement calculations, and factual inquiry.

Intoduction:

The quality of life and economic development are both affected by crime, which is a socioeconomic problem [1]. The specifics of crime prevention strategies vary with each community and culture. Training, poverty, employment, and environmental factors all have an impact on the crime rate, according to previous research on crime expectation [2]. One of the most densely populated, culturally diverse, and socioeconomically diverse major cities in Canada is Vancouver. Although the overall crime rate in Vancouver decreased by 1.5 percent in 2017, there is still a problem with high rates of car break-ins and burglaries [3]. Recently, the Vancouver Police Department (VPD) used a crime-predictive model to better anticipate crimes related to property break-ins. As a result, the city saw a 27% decrease in these types of crimes [4]. As a method of law enforcement, crime expectation relies on data and factual analysis to identify impending criminal activity [5]. Research in this area has the potential to continue in many parts of the globe. AI refers to the science of programming computers to make decisions independently of humans. Autonomous cars, speech recognition, search engines, and an improved understanding of the human genome are some of the recent applications of machine learning. It has also made it feasible to predict criminal activity based on cited data. As a controlled expectation method, arrangement takes apparent grade points into account. Numerous domains have made use of grouping, such as weather forecasting, clinical consideration, banking and finance, national security, and corporate intelligence [6]. Data collection, organisation, design recognisable evidence, expectation, and perception are the usual components of an AI-based criminal investigation. While older methods of data mining (such as affiliation analysis, grouping and expectation, bunch analysis, and anomaly examination) can only identify patterns in structured data, more recent methods can do the same with both structured and unstructured data [7]. The primary objective of this project is to develop a crime prediction model. Our analysis used two clustering algorithms, K-Nearest Neighbour (KNN) and supported choice tree, applied to the VPD crime dataset, which included more than 560,000 records collected between 2003 and 2018. Two separate approaches were used to produce the dataset:

• When a particular kind of crime happens in a certain region, the primary approach assigns an unusual number to each area and crime class.

• A two-digit number was assigned to the location and day of the week in the following technique; if the crime happened on that particular day in that specific place, the number was set to 1, and otherwise, it was put to 0.

Related Work

The problems with crime control have been addressed by several experts, who have also suggested different methods of calculating crime forecasts. The selected attributes and the dataset used as a viewpoint determine the precision of expectation. For the purpose of predicting hotspots for criminal activity in London, UK, the authors of [1] combined segment data derived from actual crime statistics with human behaviour data collected from mobile organisation mobility. Using WEKA, an open-source data mining application, and 10-fold cross-approval, an analysis was conducted in [6] to compare two clustering algorithms, Choice Tree and Naïve Bayesian. The 1990 US Enumeration, the 1990 US LEMAS review, and the 1995 FBI UCR were used separately to compile the financial, law-implementation, and criminal databases for this inquiry. In [8], several incidental elements such as the driver, weather, vehicle, and roadway conditions were taken into account while designing street accident reconstructions in Ethiopia. Three different classification algorithms, namely KNN, Naïve Bayesian, and Decision tree, were used to a dataset consisting of 18,288 accidents. Expectation exactness ranged from 79% to 81% across all three computations.An important challenge in crime prediction is efficiently and accurately analysing large crime datasets. Discovering hidden patterns in massive crime datasets is a breeze with the help of information mining. The accuracy of crime forecasting is enhanced by the enhanced efficacy and reduced errors in crime information mining methodologies. The University of Arizona-led Coplink project served as the basis for the creation of an entire information mining system in [7]. The majority of research on crime prediction is on identifying "crime areas of interest," or locations where crime rates are much higher than average. Researchers in [9] also looked at how to make area-of-interest maps and suggested region-explicit predictive models with little data using Kernel Density Evaluation (KDE) and Risk Terrain Modelling (RTM) computations. In order to predict where crimes would occur, the authors of [10] used a spatial-fleeting model that used KNN, Linear Discriminant Analysis (LDA), and factual techniques based on histograms. In order to predict where crimes will occur in Bangladesh, the authors of [11] used a crime frequency checking computation to train an Artificial Neural Network (ANN) that was further enhanced using the Gamma test. In [12], researchers in Taiwan looked at drug-related criminal data and tried to predict where the interest would arise using an AI

calculation that was data driven and based on the broken-window theory, geographical inspection, and perceptual methodologies. Using Open Street Map (OSM) and geospatial data for different types of crime in the Nova Scotia (NS) area of Canada, the authors of [13] built an AI model for crime expectation using the converse geocoding approach and a thickness-based grouping computation. For the purpose of crime prediction in Chicago, a Deep Neural Network (DNN) trained with spatial-, transient-, natural-, and joint-include depiction layers was suggested in [14]. Knowledge Discovery in Data Sets (KDD) methods, which include quantifiable demonstrating, AI, information base storage, and AI innovations, were suggested as an effective tool for crime prediction after a number of methodologies were examined in [15]. It was suggested in [16] to use crossdomain metropolitan datasets, meteorological information, markers of interest, human versatility information, and grievance information in an exchange learning system that captures worldly spatial instances. In order to simulate the dependence between the offence data and ecological parameters including segment qualities and geographical area in the territory of New South Wales (NSW), Australia, a fully probabilistic computation based on Bayesian approach was used in [17]. In [18], the relative concentration for measuring the accuracy and sufficiency of straight relapse, added drug relapse, and choice stump calculations for crime prediction in the region of Mississippi was led by WEKA.
An overview study on criminal information mining examining ANN, decision trees, rule acceptance, nearest neighbour strategies, and genetic computation was presented in [19]. To develop a robust predictive model for identifying crime trends in metropolitan locations, an approach relying on the Auto-Regressive Integrated Moving Normal model (ARIMA) was utilised in [20]. The authors of [21] suggested a method for dealing with model criminal activity in the Metro Vancouver area based on an irregular walk-based probabilistic model of spatial behaviour for actual criminals. In [22], the component of Brazilian metropolitan indicators for crime expectation was assessed using the irregular woods computation. For the purpose of developing a crime prediction solution for Chilean large urban populations, the Dempster-Shafer hypothesis of evidence, the multi-piece approach, and the upcoming strategy were used in [23]. The purpose of this study was to develop, test, and evaluate three different algorithms for crime prediction in San Francisco: KNN, Parzen windows, and Neural Networks [24].
In [25], the authors used an AI prediction model that made use of the Gradient Boosting Machine (GBM) approach to unearth hidden connections inside criminal organisations. They then used the weighted page-rank strategy as a powerful tool to cripple and eradicate these organisations. The order computations KNN and aided decision tree were used in this test based on the writing.

Data Analysis

A. Data Source

A number of datasets were first sourced from Vancouver's open data inventory. Wrongdoing and neighbourhood are the two datasets used for this endeavour. Every Sunday morning, the VPD updates the wrongdoing dataset that it has been collecting since 2003. It details the time and place of the crime as well as the nature of the infraction. Geographic information system (GIS) borders for the 22 neighbourhoods in the city are included in the neighbourhood dataset. Data analysis in this project makes use of the wrongdoing dataset, while maps are created using the neighbourhood dataset.

B. Preprocessing

The first dataset needs to be preprocessed to fill the void cells, delete pointless sections, and add a few significant highlights.

C.Statistical Analysis

The misconduct dataset's distribution as characterised by day, month, and year. As a general rule, there are roughly 3,124,000 instances of wrongdoing in Vancouver each year, 2,720,000 each month, and 90 every day. As the time intervals go longer, the dataset becomes more typical of an ordinary distribution. However, the daily chart shows an unusual maximum value of 650 events, which is thought to be an outlier - and it ultimately points to the StanleyCup rebellion on June 15, 2011.

D. Trend Analysis

From 2003 to 2013, the average monthly number of wrongdoings reduced, but in 2016, it rose again. In 2018, it fell somewhat to around 3000 instances per year, according to Fig. 3. The time-featured warming map maps indicate that the late spring and mid-month are often the most risky times of the year. On top of that, Fridays, Saturdays, and nighttime tend to have higher violation rates. The most notable values on the warmth map were around 0 hours, which should be disregarded, as all the empty data cells were blank. In terms of total occurrences, auto burglary was first, followed by wickedness. Theft from automobiles has decreased significantly in recent years, however other types of robbery have gone up.

E. Geographical Analysis

Among the many approaches to area-of-interest planning, choropleth planning is one of the most popular for describing the geographical data of wrongful episodes [26]. The density of statistical estimates or their extent is shown by shaded tones in choropleth maps. Because of this, it is easy to see places where criminal acts are more concentrated, providing insight into criminal behaviour. When it comes to plotting out criminal acts, Geographic Information System (GIS) has shown to be an extraordinary scientific tool. Police may use it for both tactical and strategic decision-making since it consolidates many geographic data sets that reveal crime hotspots into a single guide.

Proposed Methodology

Artificial intelligence (AI) is a kind of artificial intelligence that uses statistical approaches to help computers learn from their experiences [28]. Such categories as supervised, unsupervised, and assist learning help to categorise AI. Because of the concept of necessary information data and yield objectives, this research employs supervised learning. Characterisation and relapse are two types of supervised learning. Predicting an infinite number is known as relapsing, but predicting a discrete class name is known as characterisation. Predicting the types of misbehaviour in a certain region is the goal of this effort. Consequently, the chain of wrongdoing is the focus of this investigation. K-Nearest Neighbour (KNN), Backing Vector Machine (SVM), Naïve Bayesian, Decision Tree, and Ensemble Methods are just a few of the many algorithms that may be used for grouping. From a same dataset, many calculations might provide different results, each with its own set of pros and cons in terms of complexity, accuracy, and preparation time. We built our model using KNN and decision-tree computations in this study. When it comes to arranging calculations, KNN is among the easiest. If class A has the most attributes that are closest to z, then class A is assigned to the example z; otherwise, class B is assigned. The following equation is used in KNN to calculate the probability of the test being utilised for classification: KNN keeps track of all the articles that are available and uses the proximity measure to define new articles by finding the information values' closest neighbour [29]. On the other hand, decision trees are more suited to handle massive datasets with several levels including diverse nodes [18]. A decision-tree classifier may limit the number of possible decision foci while yet providing a good balance of accuracy and flexibility. The dataset is divided into smaller parts using the decision-tree-grouping model, which then uses those pieces to construct a tree structure. At each step of the computation, decision tree uses two capabilities—Gini contamination and data acquisition—to choose an inclusion that best parts the data. Gini debasement measures the probability of incorrectly ranking a random example: A

Approach 1:

The basic approach is pairing all unmitigated factors with zero and one. Highlights were created from every neighbourhood and day. Assigning "1" to the correct factor and "0" to the others is the process. The "0" factors are essentially dummy factors. This ensures that the data remains undistorted and provides the computation with more parameters to consider. The 98.9% false exactness produced by analyses using skewed data is not solid.

Approach 2:

The basic approach is pairing all unmitigated factors with zero and one. Highlights were created from every neighbourhood and day. Assigning "1" to the correct factor and "0" to the others is the process. The "0" factors are essentially dummy factors. This ensures that the data remains undistorted and provides the computation with more parameters to consider. The 98.9% false exactness produced by analyses using skewed data is not solid. AV

Conclusion

In this investigation, two distinct dataset techniques were used using misconduct data from Vancouver over the last fifteen years. Acquiring an accuracy of 39% to 44% in predicting misbehaviour was achieved using artificial intelligence predictive models KNN and enhanced decision tree. There was a little difference in the techniques and computations in terms of precision, complexity, and preparation time. By fine-tuning the computation and the data for specific uses, the accuracy of the predictions may be enhanced. Despite the model's lack of precision as a prediction model, it does serve as a foundation for further research.

References

1. H. Adel, M. Salheen, and R. Mahmoud, "Crime in relation to urban design. Case study: the greater Cairo region," Ain Shams Eng. J., vol. 7, no. 3, pp. 925-938, 2016.
2. "Overall crime rate in Vancouver went down in 2017, VPD says," CBC News, Feb. 15, 2018. [Online] Available: https://www.cbc.ca/news/canada/british-columbia/crime-rate-vancouver2017-1.4537831. [Accessed: 09- Aug- 2018].
3. J. Kerr, "Vancouver police go high tech to predict and prevent crime before it happens," Vancouver Courier, July 23, 2017. [Online] Available: https://www.vancourier.com/news/vancouver-police-go-high-tech-topredict-and-prevent-crime-before-it-happens-1.21295288. [Accessed: 09- Aug- 2018]
4. J. Han, Data mining: concepts and techniques, Morgan Kaufmann, 2012.
5. R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.
6. H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," IEEE Computer, vol. 37, no. 4, pp. 50-56, Apr. 2004.
7. T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," Proc. of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.
8. M. Al Boni and M. S. Gerber, "Area-specific crime prediction models," 15th IEEE Intl. Conf. on Mach. Learn. and Appl., Anaheim, CA, USA, Dec. 2016.
9. Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," IEEE 20th Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD), Nanchang, China, May 2016.
10. N. Mahmud, K. Ibn Zinnah, Y. Ar Rahman, and N. Ahmed, "CRIMECAST: a crime prediction and strategy direction service," IEEE 19th Intl. Conf. on Comput. and Inform. Technol., Dhaka, Bangladesh, Dec. 2016.
11. Y. L. Lin, L. C. Yu, and T. Y. Chen, "Using machine learning to assist crime prevention," IEEE 6th Intl. Congr. on Advanced Appl. Inform. (IIAIAAI), Hamamatsu, Japan, Jul. 2017.
12. F. K. Bappee, A. S. Júnior, and S. Matwin, "Predicting crime using spatial features," Can. AI 2018: Advances in Artificial Intel.-Lecture Notes in Comput. Sci., vol. 10832, pp. 367-373, Springer, Mar. 2018.
13. H. W. Kang, H. B. Kang, "Prediction of crime occurrence from multimodal data using deep learning," PLoS ONE, vol. 12, no. 4, Apr. 2017.
14. V. Grover, R. Adderley, and M. Bramer, "Review of current crime prediction techniques," Intl. Conf. on Innovative Techn. and Appl. of Artificial Intel., pp. 233-237, Springer, London, 2007.
15. X. Zhao and J. Tang, "Exploring transfer learning for crime prediction," IEEE Intl. Conf. on Data Mining Workshop (ICDMW), New Orleans, LA, USA, Nov. 2017.

16. R. Marchant, S. Haan, G. Clancey, and S. Cripps, "Applying machine learning to criminology: semiparametric spatialdemographic Bayesian regression," Security Inform., vol. 7, no. 1, Dec. 2018.
17. L. McClendon and N. Meghanathan, "Using machine learning algorithms to analyze crime data," Mach. Learn. and Appl.: an Intl. J. (MLAIJ), vol.2, no.1, Mar. 2015.
18. S. Prabakaran and S. Mitra, "Survey of analysis of crime detection techniques using data mining and machine learning," Nat. Conf. on Math. Techn. and its Appl. (NCMTA 2018), IOP J. of Physics: Conf. Series, vol. 1000, 2018.
19. E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models," IEEE 14th Intl. Conf. on Dependable, Auton. and Secure Comput., Auckland, New Zealand, Aug. 2016.
20. M. A. Tayebi, U. Glässer, and P. L. Brantingham, "Learning where to inspect: location learning for crime prediction," IEEE Intl. Conf. on Intel. and Security Inform. (ISI), Baltimore, MD, USA, May 2015,.
21. L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues, "Crime prediction through urban metrics and statistical learning," Physica A, vol. 505, pp. 435-443, 2018.
22. N. Baloian, E. Bassaletti, M. Fernández, O. Figueroa, P. Fuentes. R. Manasevich, M. Orchard. S. Peñafiel, J. A. Pino, and M. Vergara, "Crime prediction using patterns and context," IEEE 21st Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD), Wellington, New Zealand, Apr. 2017.
23. M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Sep. 2016.
24. E. Budur, S. Lee, and V. S. Kong, "Structural analysis of criminal network and predicting hidden links using machine learning," arXiv:1507.05739, Sep. 2015.
25. G. Zhou, J. Lin and W. Zheng, "A web-based geographical information system for crime mapping and decision support," IEEE Intl. Conf. on Comput. Problem-Solving (ICCP), Leshan, China, Oct. 2012.
26. S. Chainey and J. Ratcliffe, GIS and crime mapping, John Wiley & Sons, Incorporated, 2015.
27. M. Stamp, Introduction to machine learning with applications in information security, Chapman and Hall/CRC, New York, 2017.
28. S. Sivaranjani, S. Sivakumari, and M. Aasha, "Crime prediction and forecasting in Tamilnadu using clustering approaches," IEEE Intl. Conf. on Emerg. Technol. Trends (ICETT), Kollam, India, Oct. 2016.
29. N. Shama, A machine learning approach to predict crime using time and location data, B.Sc. Thesis, BRAC University, 2017.