International Journal of Scientific Research & Engineering Trends
Volume 8, Issue 5, Sep-Oct-2022, ISSN (Online): 2395-566X

A Review of Crime Detection Using Machine Learning

Sameeksha Bhati, Assistant Professor Priyanshu Dhameniya
Department of Computer Science
Astral Institute of Technology and Research,
Indore,MP,India

Abstract-As a societal and economic issue, crime has a negative impact on people's standard of living and the health of the economy [1]. The particulars of criminal behavior vary greatly from one group or community to the next. The crime rate can be predicted, at least in part, by looking at socioeconomic indicators such as levels of education, poverty, unemployment, and weather. Vancouver, one of Canada's most populous and diverse urban centers, is home to people of many different cultural backgrounds and ancestries. There was a 1.5% decrease in 2017's overall crime rate in Vancouver, although a persistent problem remains with car theft. The residential burglary rate in Vancouver dropped by 27 percent after the Vancouver Police Department (VPD) used a crimepredictive model to predict such crimes. Predicting criminal activity is a tool used by law enforcement that relies on data and statistical analysis to pinpoint potential hotspots. Research in this area has continued in several countries.

Keywords **– Crime Detection,Machine learning,ML Clasifiers**

I. INTRODUCTION

People commit crimes because they have the means to do so, and these criminal acts are in turn driven by a variety of social, economic, and political dynamics. When anything is illegal, it is either done or not done in defiance of a prohibition or order from the law, respectively. The punishment for crimes may be life in prison, probation, a fine, or a restraining order. In addition, they may be subject to fines or other penalties. Because they occurred frequently, these crimes were labeled "serial". [1, 2].

Crime Detection

Authorities tasked with maintaining public safety, such as police patrols, work tirelessly to identify criminals and prevent such crimes from occurring in their communities. To maintain public security and peace of mind, the police force takes preventative and investigative crime-detection procedures. There is a wide variety of crimes, and not all of them are documented in the system's data [3]. Crime takes on varying manifestations in different parts of the world as a result of the varied ways in which local cultures are incorporated into mainstream society.

When authorities lack the necessary resources, it is difficult to track down crime scenes and understand criminal behavior. One of the primary goals of crime analysis is to reassure the public that criminal activity is under control and poses no threat to public safety. Preventing crime may be a fundamental problem that has plagued humanity for aeons. It's generally understood to be a collection of methods for minimizing illegal acts and maximizing the efficiency with which available resources are put to use. Law enforcement organizations should increase their use of criminological theories and crime analysis tools in order to develop effective policies and enhance prevention strategies. Evaluation of information using conventional methods is laborious, time-consuming, and dependent on specialized knowledge in areas such as finance, economics, business, and the law. To reduce criminal activity at a certain address and time, an automated procedure must be put into place. Many studies have been conducted in an attempt to pinpoint the most dangerous areas. Crime evaluation mapping can be carried out quickly and accurately with the use of a geographic information system. [4].

Crime Analysis

Quantitative and qualitative analyses of crime and law enforcement statistics, as well as consideration of socioeconomic and geographical aspects, are examined in these studies. Crime analysts use a wide variety of data and methods, tailored to the specific situation and analyst. Crime analysis refers to a series of analytical procedures that help law enforcement agencies with crime prevention by, among other things, examining historical crime data, forecasting future trends, and planning resource allocations. Numerous police departments rely on it to aid in their efforts to reduce crime and understand what went wrong [4, 5]. In Figure 1.1, we see that there are four main subfields of criminological analysis.

© 2022 IJSRET

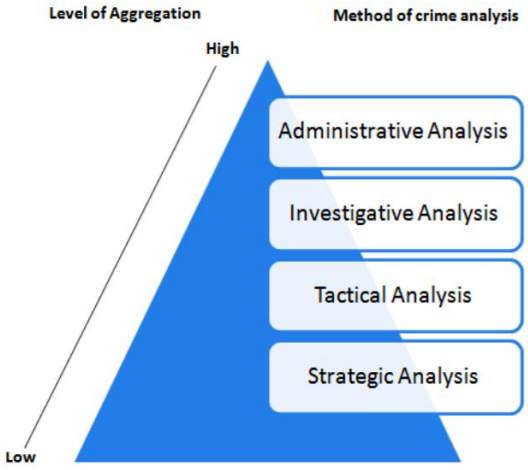

Figure 1.1: Crime Analysis Model

II.DATA MINING IN CRIME DETECTION

Violent crime is deplorable because it lowers community standards. Criminal behavior patterns can be identified and cases can be closed more quickly with the use of data mining tools. The data mining methods used in the process of crime clustering are crucial [10]. Geocrime refers to a cluster of crimes that occur in close proximity to one another within a defined geographic area. The crime's overlay on the map can be visualized through the use of geo-spatial plotting. People can use the highly populated crime cluster to better pinpoint the location of the criminal activity. In data mining, clustering is used to spot a consistent criminal behavior pattern or a string of related crimes. The atrocities committed by the DC sniper, serial rapists, and serial murders are among the most common. One person or a small cell could be responsible for these acts.

When it comes to figuring out how to group crimes together, the crime dataset is crucial. Some of the information in the dataset is private while others are available to the public. However, statistics on drug-related crimes or cases involving minors are often scarcer. Similarly, while warning the public about sex offenders, the victim's identity is often withheld so as to protect them. If a data miner or analyst wants to stay on the right side of the law during the data mining modeling process, they'll need to address concerns about what information can be made publicly available and what should be kept private. Information such as the nature of the crime, its date and time, its location, and other similar details can be found in these reports. And then there's the details about the culprit (whether known or unknown), the victim, and the witnesses. The MO and a description of the crime are typically written down as well. Free-form text is used for the vast majority of police and detective observations since they do not easily fit into predetermined question sets with checkboxes. Whereas databases normally store the first two types of data as numeric, character, or date fields of table, the third type of data is often saved as free text.

II. LITERATURE REVIEW

Ashok kumarPalanivinayagam et al. [1],One of the primary uses of AI in examination is in criminal detection. The key to developing a healthy society is recognizing and reducing crime rates. Information is gathered and investigated using Huge Data techniques, with the goal of identifying the prime credits and necessary ingredients that led to the growth of crime hotspots. Most of the time, traditional wrongdoing detection and AI-based calculations fail to properly predict wrongdoing designs because they are unable to generate essential prime credits from the wrongdoing dataset. In order to improve the precision of the AI calculation at hand, this study is geared on isolating the major attributes, such as time zones, wrongdoing likelihood, and wrongdoing regions, and conducting weakness investigation. We implemented the recommended method by using two industry-standard datasets. The findings demonstrate that the proposed emphasize age method enabled more AI models to be presented. When used to the Nave Bayes computation while analyzing the San Francisco dataset, the proposed approach achieved the highest accuracy of 97.5%.

O. Vynokurova et al. [2],The problem of extortion id is becoming more severe as technology advances in tandem with it. From one angle, the rise in the volume of electronic transactions across a variety of corporate contexts, and from another, the advancements in technology and programming, have contributed to a functional rise in electronic crime. In the paper, a hybrid AI setup is provided for handling tasks related to anomaly detection. Both the oddities detection subsystem (with individual study in mind) and the oddity type translation subsystem make up this hybrid framework (in light of regulated framework). When information needs are attended to in a continual fashion, the proposed half-and-half framework can whiz through them quickly. During the planning of the abnormality recognition problem using real-world data streams, the proposed approach was confirmed to be effective.

E. E. Eryilmaz et al. [3],Modern letters are sent via the internet via electronic mail. Various types of data, including as archives, images, audio, and recordings, can be attached to messages and sent to the recipient's computer. Because of their low cost and ease of use, billions of messages are sent out every day. Since email is the most popular technological tool for both personal and professional communication, it is clear to see why it has become so popular: it is an efficient means of communication that helps save both time and money. Because of their practicality and low cost, messages are widely used by individuals and networks for

© 2022 IJSRET

dissemination, advertising, and phishing. In order to get their jobs done, people or groups may send random, meaningless emails to people they will never meet. These messages cause real harm to web users in terms of both money and reputation, and they slow down traffic. Email spam is a distribution tactic that is sent to a recipient without his knowledge or permission and is typically employed for malicious or other inappropriate reasons. The goal of spammers is to influence online consumers to purchase goods and services, whether they are legal or not. Since spammers constantly update their methods, new artificial intelligence-based spam identification technologies have emerged. This article discusses a study that used seven distinct artificial intelligence methods to detect spam in 800 datasets of Turkish email. The best result is obtained from the Sequential Minimal Optimization (SMO) computation in the developed method when the component determination is done with the chi-square test. Multi-Layer Perceptron (MLP) calculations provide the best results when used for element determination during the data acquisition approach. The SMO and MLP F-measure execution values are 0.985 and 0.984, respectively.

A. Palanivinayagam et al. [4],In order to understand what is being said to us, several new methods of discovery are proposed for the coming year. We present a sophisticated iterative bunching calculation that makes use of the interpreted text and reduces error in this study. The suggested method consists of three steps, (1) the obscure word likelihood task, (2) multi-likelihood standardization, and (3) likelihood sifting, all of which are carried out across several emphasis levels. Each cycle, generally speaking, takes in the obscure words from the previous cycle's emphasis and distributes likelihood to the obscure words based on the transitory results obtained in the previous cycle's emphasis. Repeat until no more cryptic phrases remain. Next, we'll see how taking into account the probability of adjacent words can help standardize the multiple probabilities assigned to a single word. The most recent development is the elimination of possibilities below the rim, which ensures a reduction in noise levels. Using authentic benchmark datasets, we quantify the characteristics of bunching. The outcomes demonstrate that, in comparison to alternative grouping calculations, our enhanced calculation produces more accurate bunching.

A. R. Javed et al. [5],Recently, technological advancements have facilitated an increasingly important function in the dissemination and exchange of knowledge. Convincing messages, from proclamation to coercion, are spread with the help of visual depictions and first-hand experiences of events. The primary focus of the current sophisticated network for investigating misconduct is on protecting, recovering, and analyzing data already stored in computers. Physically collecting, processing, and analyzing all media content would be a time-consuming and taxing task for humans. In this study, we present a creative approach to investigating the assumed machine, which solves a persistent problem for an agent. Our method involves retrieving all image data from the supposing machine at the byte level, performing fast and accurate article location using a deep learning-based calculation, and presenting an incontestable level representation of images with a questionable item or remarkable item that can be introduced as evidence. Our approach involves praising images containing questionable items. This review confirms the importance of automated article recognition in digital crime scene investigation, based on its efficacy and efficiency of implementation.

B. S. Aldossari et al. [6],Crime has an evident impact on society and has a significant impact on individual victims' minds, concerns, and cognitive abilities. Therefore, law enforcement authorities continue to monitor the areas within their jurisdiction in order to detect any suspicious behavior, which makes them more alert and improves their ability to prevent crimes. The purpose of this analysis is to provide context for future research into creating an AI model with the ability to think about the classification of wrongdoing in a certain domain. Choice Tree and Naive Bayes are utilized on a dataset that was not included in the Chicago Police Department's CLEAR. CLEAR did not incorporate this data collection. The ability to foresee and prevent prospective infractions is a major benefit of these processes for law enforcement authorities. Two calculations were utilized to analyze the nine most notable features of the dataset we selected. The Decision Tree outperformed the Naive Bayes method in a head-to-head comparison.

A. Chowdhary et al. [7], Extraction of useful information from images or image arrangements is a large subfield of computer vision. One such area where much of research is being done is in the area of human action recognition. This approach has vast practical implications for many types of studies and practice. A convolutional neural architecture with transfer learning and a long transient memory model are proposed here as part of a two-model technique. Each video's element vectors are accumulated using a CNN network, and then the LSTM network is used to arrange the scenes in chronological order. Bench presses, horse rides, basketball dunks, and so on are all examples of standard exercises. The proposed calculation achieved a high degree of precision of 94.2%.

P. Ashok Kumar et al. [8], The location of a business is the single most important factor to think about before launching it, because more people will visit a business

© 2022 IJSRET

in a prime location. Web crawlers help customers find local businesses before making a trip. If a customer needs to purchase a gemstone, for instance, he can look it up in a web index to locate the nearest gemstone retailer. He is looking for the optimal location to open a new gem store, preferably one where there is neither a competing gem dealer nor a dearth of potential customers. In this research, we offer a method for determining the optimal location for launching a business in an area with high levels of both positive and negative attention (and not many stock). Our new approach outperforms the standard kNN computation on a number of metrics, including average support duration and client business share, which we test on benchmark datasets.

A. R. Javed et al. [9],As technology improved and more people started using their phones for things like banking on the go, they became more vulnerable to malicious assaults. When using a smartphone's sensitive keyboard, a user's fingers generate vibrations that can be used in side-channel attacks to infer which keys are being pressed. We develop and evaluate AlphaLogger, an Android app that can figure out the sequence of keys being pressed on a nuanced keyboard. AlphaLogger operates invisibly, collecting data from the accelerometer, gyroscope, and magnetometer in the mobile device at a rate of 10Hz/sec, allowing it to accurately infer the keystrokes being composed on the sensitive console of any other programs still active in the foreground. We offer a study of the presentation sensor combinations. Extensive testing demonstrates that the accelerometer, spinner, and magnetometer may be used to accurately predict key presses to within 9.2 percent.

3.Machine Learning Approach

Words like "machine learning" and "artificial intelligence" have a tremendous impact on the numbers. That AI keeps become more well-known is a direct effect of how much we rely on it in our daily lives. It's now commonplace to find them built into all sorts of machinery and tools. Nevertheless, caution is still exercised. Let's look into ML's brief history to dispel such misconceptions. Machine learning (ML) can be defined as a cycle of providing data to computer systems in order to improve those systems' ability to measure and do the action in question in the future without being explicitly adjusted or cared for using similar or additional data. If computers are given the ability to reason, they will become smarter and easier to operate. Their value to society will skyrocket, making them an essential commodity. Nearly every branch of knowledge can benefit from ML.

4.Supervised ML Clasifiers

Classifiers trained with supervised ML can be broken down into distinct categories. Some examples of these include the belief-based Bayes, discriminant analysis (both linear and quadratic), summed direct models, stochastic inclination plunge, support vector machines, linear support vector classifiers, decision trees, neural organization models, closest neighbors, and outfit approaches. The tactics used by the troupe help weaker pupils come together to form stronger ones. These predictive models are built with the intention of increasing the overall level of accuracy. There are two methods to achieve this. The first approach is to employ highlight designing, while the second method is to make use of boosting computations. Those impressions of preparation that end up being misclassified are the target of enhancement calculations. AdaBoost, CatBoost, LightGBM, XGBoost, and gradient boosting are five of the most popular boosting techniques.

3. CONCLUSION

Most approaches for detecting crimes and analyzing hotspots for criminal activity employ various processing techniques on the raw data to arrive at a reliable conclusion. However, there are numerous issues that cause subpar results. The development of inaccurate synthetic minority samples is a major concern in the literature, as is the class imbalance problem and multi-objective problems caused by the different classes. Graph based clustering algorithms require full processing of the graph regardless of how tiny the changes are, making their application more difficult. It is difficult to maintain a high level of accuracy when classifying crimes that are similar to one another.

REFERENCES

[1] AshokkumarPalanivinayagam, Siva Shankar Gopal, Sweta Bhattacharya, Noble Anumbe, EbukaIbeke and CresantusBiamba, “An Optimized Machine Learning and Big Data Approach to Crime Detection”, Wireless Communications and Mobile Computing, Hindawi, 2021.

[2] O. Vynokurova, D. Peleshko, O. Bondarenko, V. Ilyasov,V. Serzhantov, and M. Peleshko, “Hybrid machine learningsystem for solving fraud detection tasks,” in 2020 IEEE ThirdInternational Conference on Data Stream Mining Processing(DSMP), pp. 1–5, Lviv, Ukraine, 2020.

[3] E. E. Eryilmaz, D. O. Ahin, and E. Kl, “Machine learning basedspam e-mail detection system for Turkish,” in 2020 5th InternationalConference on Computer Science and Engineering(UBMK), pp. 7–12, Diyarbakir, Turkey, 2020.

[4] A. Palanivinayagam and S. Nagarajan, “An optimized iterativeclustering framework for recognizing speech,” InternationalJournal of Speech Technology, vol. 23, no. 4, pp. 767–777, 2020.

[5] A. R. Javed and Z. Jalil, “Byte-level object identification forforensic investigation of digital images,” in 2020 InternationalConference on Cyber Warfare and Security (ICCWS), Islamabad,Pakistan, 2020.

© 2022 IJSRET

[6] B. S. Aldossari, F. M. Alqahtani, N. S. Alshahrani et al., "Acomparative study of decision tree and naive bayesmachinelearning model for crime category prediction in Chicago," inProceedings of 2020 the 6th International Conference on Computingand Data Engineering, ser. ICCDE 2020, p. 3438, NewYork, NY, USA, 2020.

[7] A. Chowdhary and B. Rudra, "Video surveillance for the crimedetection using features," in Advances in Intelligent Systemsand Computing Advanced Machine Learning Technologiesand Applications, pp. 61–71, Cairo, Egypt, 2020.

[8] P. Ashok Kumar, G. Shiva Shankar, P. K. R. Maddikunta, T. R.Gadekallu, A. Al-Ahmari, and M. H. Abidi, "Location basedbusiness recommendation using spatial demand," Sustainability,vol. 12, no. 10, p. 4124, 2020.

[9] A. R. Javed, M. O. Beg, M. Asim, T. Baker, and A. H. Al-Bayatti, "AlphaLogger: detecting motion-based side-channelattack using smartphone keystrokes," Journal of AmbientIntelligence and Humanized Computing, vol. 1, 2020.

[10] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna et al., "Analysisof dimensionality reduction techniques on big data," IEEEAccess, vol. 8, pp. 54776–54788, 2020.

[11] P. Ashokkumar and S. Don, "Link-based clustering algorithmfor clustering web documents," Journal of Testing and Evaluation,vol. 47, no. 6, p. 20180497, 2019.

[12] R. Jain, A. Nayyar, and S. Bachhety, "Factex: a practicalapproach to crime detection," in Data Management, Analyticsand Innovation Advances in Intelligent Systems and Computing,p. 503516, Pune, India, 2019.

© 2022 IJSRET