فصلنامهٔ علمی ـ پژوهشی رفاه اجتماعی، سال بیست و دوم، بهار ۱۴۰۱، شماره ۸۴

Social Welfare Quarterly, Vol 22, Spring 2022, No 84

استخراج الگوهای جرائم مواد مخدر و شناسایی افراد در معرض خطر با استفاده از تکنیکهای داده‌کاوی

۱. احمد بختیاری شهری
دکتر فناوری اطلاعات، گروه فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران.(نویسنده مسئول)، <Bakhtiyari@ece.usb.ac.ir>

۲. سمیرا نوفرستی
دکتر نرم‌افزار، گروه فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران.

۳. نصرت افتخاری
کارشناس ارشد مهندسی تجارت الکترونیک، دانشکده مهندسی برق و کامپیوتر، دانشگاه بین‌المللی چابهار، چابهار، ایران

۴. نادیا جهانتیغ
کارشناس ارشد مهندسی کامپیوتر، گروه فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه سیستان و بلوچستان، زاهدان، ایران.

واژه‌های کلیدی:
جرائم مواد مخدر، داده‌کاوی، طبقه‌بند نزدیک‌ترین همسایه، الگوهای جرم

تاریخ دریافت: *۱۴۰۰/۰۳/۰۶*
تاریخ پذیرش: *۱۴۰۰/۱۰/۲۱*

مقدمه: *رشد فناوری اطلاعات در سازمانها، منبع عظیمی از داده‌های ذخیره‌شده در حوزه جرائم مرتبط با مواد مخدر را فراهم آورده است. تحلیل این داده‌ها و کشف الگوهای پنهان موجود در آن به کمک داده‌کاوی می‌تواند به کشف و پیشگیری از وقوع جرائم در این حوزه کمک کند. هدف این مقاله به‌کارگیری تکنیکهای داده‌کاوی برای شناسایی افراد مستعد به قاچاق مواد مخدر در استان سیستان و بلوچستان و نیز کشف الگوهای جرم است.*

روش: *پژوهش حاضر روی داده‌های ۴۶۷ مجرم حوزه مواد مخدر در استان سیستان و بلوچستان که در طی سالهای ۱۳۹۲ الی ۱۳۹۹ مرتکب جرم قاچاق مواد مخدر شده‌اند با نمونه‌گیری در دسترس انجام گرفته است. برای انجام این تحقیق از متدولوژی استاندارد CRISP-DM و الگوریتمهای طبقه‌بندی ماشین بردار پشتیبان، بیزین ساده، رگرسیون لجستیک، درخت تصمیم و نزدیک‌ترین همسایه و برای استخراج الگوهای جرائم از الگوریتم الگوکاوی اپریوری استفاده شده است.*

یافته‌ها: *الگوریتم الگوکاوی بالغ بر ۲۰ الگوی جرم با دقت بالای ۸۰ درصد استخراج کرده است. به علاوه در میان الگوریتمهای طبقه‌بندی، طبقه‌بند نزدیک‌ترین همسایه قادر است با دقت ۸۴ درصد افراد در معرض خطر را شناسایی کند.*

بحث: *با به‌کارگیری مدل ساخته‌شده با این الگوریتم می‌توان سامانه‌ای برای شناسایی افراد مستعد به قاچاق مواد مخدر طراحی کرد. نتایج حاصل از پیش‌بینیهای انجام گرفته توسط سامانه مذکور و کشف الگوهای پنهان موجود در داده‌ها می‌توانند کمک شایانی به پلیس، دستگاههای قضایی و مددکارهای اجتماعی در شناسایی افراد در معرض خطر و کاهش جرائم مرتبط با قاچاق مواد مخدر کند.*

Social Welfare Quarterly, Vol 22, Spring 2022, No 84

Extraction of Drug Crime Patterns and Identifying People at Risk Using Data Mining Techniques

1-**Ahmad Bakhtiyari Shahri**
Ph.D. in Information Technology, Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Zahedan, Iran, (Corresponding Author)
<bakhtiyari@ece.usb.ac.ir>

2- Samira Noferesti
Ph.D. in Software, Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Sistan and Baluchistan, Zahedan, Iran.

3- Nosrat Eftekhari
M.A. in Electronic Commerce Engineering, Faculty of Electrical and Computer Engineering, Chabahar International University, Chabahar, Iran

4- Nadia Jahantigh
M.A. in Computer Engineering, Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Sistan and Baluchistan, Zahedan, Iran

Keywords:
Drug offenses, Data mining, Support vector machine (SVM), Crime patterns, Crime prediction
Received: 2021/05/27
Accepted: 2022/01/11

Introduction: Introduction: In recent years, technology advancement and the growth of information technology in organizations have provided a huge source of data stored in the field of drug-related crimes. Analyzing these data and discovering hidden patterns in it can help detect and prevent the occurrence of crimes in this area. This paper aimed to identify the susceptible people to drug trafficking in Sistan and Baluchestan province and discover patterns of crime employing data mining techniques.

Method: The present study was conducted on data of 467 drug criminals in Sistan and Baluchestan province who committed drug trafficking in the years 2011 to 2020 by available sampling. CRISP-DM methodology was used to build a prediction model. Also, Support Vector Machine (SVM), Naïve Bayes, Logistic Regression and Decision Trees have been used to predict people at risk and Apriori Algorithm has been used to extract crime patterns.
Findings: The pattern mining algorithm extracted over 20 crime patterns with a precision of over 80%. The results of the evaluations show that the IBK classifier can accurately identify 84 % of the people at risk.

Discussion: A system for identifying people susceptible to drug trafficking can be designed using the model made by the IBK classifier. In addition, the results of the predictions by the above mentioned system, as well as the extracted hidden patterns can help police, judiciary, and social workers to identify people at risk and reduce drug-related crimes.

Extended Abstract

Introduction: Drug trafficking is a major problem in many countries, which has irreparable social and economic harm for families and, consequently, for the countries. Sistan and Baluchestan province is one of the smuggling transit centers in Iran due to its geographical location. Therefore, securing the province from crimes by determining the factors affecting drug trafficking, identifying people at risk and providing appropriate strategies for reducing drug crimes should be one of the priorities of the relevant authorities in Sistan and Baluchestan. Studies have shown that many unwanted people are absorbed to these deadly substances for livelihood or curiosity; therefore, one of the methods used to solve this social problem is to identify people susceptible to activities in the area of drug trafficking. The rapid growth of information technology and information systems development has driven organizations into a sea of data. Analyzing these data and discovering the hidden knowledge in them can help managers and authorities of relevant organizations to make more accurate and faster decision-making. In this regard, with the expansion of online crime information systems and the storage of criminals' data in databases, there are numerous studies around the world about the use of data mining for crime detection, crime prediction, and prevention. However, the studies carried out in the field of drug-related crimes are few. Despite the valuable and high-quality data on crimes and offenders, studies on the prediction of crime in our country have not been generalized to the drugs field and are mainly limited to the detection of theft related crimes. This paper aimed to identify the susceptible people to drug trafficking in Sistan and Baluchestan province and discover patterns of crime using data mining techniques

Method: In this paper secondary data resources assist in providing the relevance of crime detection using data mining techniques. An exploration through on-line search has been carried out among the various search databases. In order to use data mining tools and algorithms, a computer-readable dataset is required. In this regard, the files of 467 drug criminals in Sistan and Balouchestan province, involved in drugs trafficking, were collected in the years 2011 to 2020 by available sampling. Moreover, to protect the privacy of individuals, the identity of criminals including name, surname, father's name, case number, and birth certificate number were removed from the database. Cross Industry Standard Process for Data Mining methodology

and Support Vector Machine, Naïve Bayes, Logistic Regression and Decision Tree classifiers have been used to do this research and Apriori Algorithm has been used to extract crime patterns. WEKA data mining software has been used to analyze data.

Findings: This paper conducts accurate analyses of stored data on drug offenses using data mining techniques and help managers in identifying people susceptible to drug trafficking in Sistan and Baluchestan province by extracting hidden patterns in the data. Given the importance of drug trafficking in Sistan and Baluchestan province, providing solutions to identify patterns indicating the reasons of crime, as well as identifying people at risk, are indispensable. Hence, in this paper, a system to identify people susceptible to drug trafficking (i.e., those susceptible to repeat offending) was designed by analyzing data on drug offenders using classification algorithms. The pattern mining algorithm extracted over 20 crime patterns with a precision of over 80%. The results of the evaluations show that the SVM classifier can identify people at risk of drug trafficking with the accuracy of 84%. Also, crime patterns are extracted using the association rule mining algorithm.

Identified patterns show that married people from Zahedan who have been in bad financial situation, and married people aged between 31 to 40 and also married people who do opium trafficking to earn money, are committing crime again. Improving the livelihood of the people of this province, creating job opportunities especially for married people, and addressing the problems and concerns of people aged 31 to 40 should be prioritized to reduce the crimes related to drug trafficking. In addition, 84% of people who smuggle opium and were born in Zabol have 23 to 30 years old. Patterns show that people aged 31 to 40 with multiple drug traffickers have a bad financial situation and suggests that one of the reasons for drug trafficking in the age group of 31 to 40, who are mostly married and put the burden on a family, have a bad financial situation. Therefore, addressing the livelihood of the people and creating job opportunities by the authorities can be helpful.

In contrast, self-employed people who have a good financial situation, mainly smuggle opium; some people with a good financial situation are also committing drug-related offenses. The main reasons for this are unawareness of the consequences of drug trafficking, recreation or friendship with bad persons. In this regard, solutions such as informing and educating people through the creation of television programs, as well as

the broadcasting of leaflets and posters, especially in public gatherings, and the spread of healthy recreational activities particularly in the youth age group can be effective.

Discussion: Discovered patterns in this study can help the relevant authorities to identify people at risk and identify the reasons of crime. The system for identifying people susceptible to drug trafficking can be used as a police decision-making system to reduce the scope of police investigation and detection and identify offenders with spending less time and money. Also, the extracted crime patterns help the police and social workers to identify the main reasons for committing a crime and provide preventive laws to reduce drug offenses. Although, this paper analyzed drug-related data, the proposed method can be used for other types of crime. There were many problems during this research, such as limitations to collect data. Electronically saving criminals information instead of using the paper files can greatly reduce the problems of data aggregation and preparation, as well as the human error caused by manual recording of information, and it can also save time. In addition, better and more accurate patterns, including spatial patterns of crime can be extracted by having more information like a comprehensive spatial database, which includes the demographic characteristics of offenders and the characteristics of the crime and the exact location of the crime. In the future, other data-mining algorithms can be used to predict people susceptible to drug trafficking. In addition, by analyzing the extracted patterns by drug experts, it is possible to identify the main reasons of crime and those susceptible to crime, and to identify ways to reduce crime. Also, by psychologically analyzing patterns by social workers and other relevant professionals, committing similar crimes can be largely prevented.

Ethical Considerations

Authors' contributions

All authors contributed in producing of the research.

Funding

The present study did not have any sponsors.

Conflicts of interest

The authors declared no conflict of interest.

Acknowledgments

In this article, all rights relating to references are cited and resources are carefully listed.

مقدمه

یکی از اصلی‌ترین معضلات موجود در مناطق مرزی به‌ویژه در استان سیستان و بلوچستان، قاچاق مواد مخدر است که آسیبهای اجتماعی جبران‌ناپذیری را برای خانواده‌ها و به‌تبع آن برای کشور دارد. استان سیستان و بلوچستان به دلیل موقعیت جغرافیایی و هم‌جوار بودن با دو کشور تولیدکننده مواد مخدر، شرایط اقتصادی نامناسب نظیر نرخ بیکاری بالا، فقر و محرومیت، گستردگی و تنوع محیط جغرافیایی مرزهای شرقی و کمبود تجهیزات مراقبت از مرز یکی از مراکز ترانزیت قاچاق این بلای خانمان‌سوز است؛ به‌طوری‌که سالانه حدود ۱۲۰ تن مواد در این استان کشف شده و تعداد مجرمان این حوزه تقریباً ٤۰۰۰ نفر است (رحیمی و کاویان، ۲۰۱۵). ازاین‌رو بررسی عوامل مؤثر بر قاچاق مواد مخدر، شناسایی افراد در معرض خطر و ارائه راهکارهای مناسب برای کاهش جرائم مواد مخدر باید یکی از اولویتهای مسئولان مربوطه در استان سیستان و بلوچستان باشد (رحیمی و کاویان، ۲۰۱۵).

مطالعات نشان داده است بسیاری از افراد ناخواسته، به دلیل فقر، محرومیت و نبود فرصت شغلی برای امرارمعاش و یا از روی کنجکاوی طعمه این سودای مرگ‌آور می‌شوند؛ لذا یکی از روشهای مورداستفاده برای رفع این معضل اجتماعی شناسایی افراد مستعد به فعالیت در حوزه قاچاق مواد مخدر است (بختیاری و افتخاری، ۲۰۱۹).

یکی از روشهای شناسایی افراد مستعد به فعالیت در حوزه قاچاق مواد مخدر که در سالهای اخیر موردتوجه قرار گرفته است، داده‌کاوی است. سرعت رو به رشد فناوری اطلاعات و توسعه سیستمهای اطلاعاتی، سازمانها را در دریایی از داده‌ها غرق کرده است. تحلیل این داده‌ها و کشف دانش نهفته در آنها می‌تواند مدیریت و مسئولان سازمانهای مربوطه را در انجام تصمیم‌گیریهای دقیق‌تر و سریع‌تر یاری کند. در این راستا، با گسترش سامانه‌های اطلاعاتی برخط ثبت جرائم و ذخیره اطلاعات مجرمان در بانکهای اطلاعاتی،

تحقیقات متعددی در سطح جهان در خصوص استفاده از تکنیکهای داده‌کاوی برای شناسایی جرائم، پیش‌بینی جرائم و پیشگیری از وقوع آنها انجام گرفته است (تایل و همکاران، ۲۰۱۵).

با این وجود مطالعات انجام گرفته در کشور ما به دلیل عدم دستیابی به داده‌های قضایی مواد مخدر اندک و عمدتاً محدود به شناسایی و کشف جرائم مرتبط با سرقت است (اسکندری، علیزاده و کاظمی، ۲۰۱۱).

پیش‌بینی جرم شاخه‌ای از آینده‌پژوهی است که به مطالعه آینده‌های فرضی به منظور کسب آمادگی برای مقابله با آن می‌پردازد؛ که کاربردهای فراوانی در تصمیم‌گیریهای قضایی ازجمله اعطای آزادی مشروط به فرد محکوم دارد. علاوه بر این، پیش‌بینی جرم می‌تواند به شناسایی مجرمان کم‌خطر پرداخته و از این طریق به تنظیم سیاستهای کیفری متناسب در مورد محکومان و زندانیان و به‌تبع آن برقراری عدالت در بین مجرمان کمک کند(غلامی و برزگر، ۲۰۱۸).

با وجود داده‌های با ارزش و باکیفیت درباره جرائم و مجرمین مواد مخدر، تاکنون مطالعات انجام گرفته در زمینه پیش‌بینی ارتکاب جرم در کشور ما به حوزه مواد مخدر تعمیم نیافته است (بختیاری و افتخاری، ۲۰۱۷). لذا این مقاله سعی دارد با استفاده از تکنیکهای داده‌کاوی، تحلیلهای دقیقی بر روی داده‌های ذخیره‌شده درباره جرائم مربوط به مواد مخدر انجام دهد و با استخراج الگوهای پنهان موجود در داده‌ها، مدیران را در شناسایی افراد مستعد به قاچاق مواد مخدر در استان سیستان و بلوچستان یاری کند.

سامانه شناسایی افراد مستعد به قاچاق مواد مخدر می‌تواند به عنوان یک سیستم تصمیم‌یار پلیس برای کوچک‌تر کردن دایره تحقیقات و بررسیهای پلیس و شناسایی مجرمان با صرف زمان و هزینه کمتر استفاده شود. همچنین الگوهای جرم استخراج‌شده به پلیس و مددکارهای اجتماعی برای شناسایی دلایل اصلی ارتکاب جرم وضع قوانین پیشگیرانه برای کاهش جرائم حوزه مواد مخدر کمک می‌کنند.

مبانی نظری و پیشینه پژوهش

چارچوب این پژوهش بر اساس مفاهیم ماده مخدر، اعتیاد، قاچاق مواد مخدر و داده‌کاوی استوار است که در ادامه موردبحث قرار می‌گیرند.

ماده مخدر، ماده‌ای طبیعی یا شیمیایی است که از طریق آثاری که در احساسات و ذهن انسان می‌گذارد، باعث آثار مخربی در جسم و رفتار می‌شود. اغلب منظور از مواد مخدر موادی مانند هروئین یا تریاک است که منع قانونی دارند. سازمان بهداشت جهانی اعتیاد را حالت سرمستی مزمنی که براثر استفاده مکرر از مواد مخدر در فرد و جامعه اختلال ایجاد می‌کند تعریف می‌کند. خصوصیات بارز اعتیاد شامل میل شدید و غیرقابل‌کنترل برای به دست آوردن مواد به هر قیمتی، ازدیاد مقدار استفاده از آن به نحو تصادفی و اتکای شدید روانی و گاهی جسمانی به استفاده از آن مواد است (ابراهیمی و بافندره، ۲۰۱۷). در حقوق ایران، جرائم صادر کردن، واردکردن، حمل کردن، ارسال و ترانزیت مواد مخدر از مصادیق جرائم قاچاق مواد مخدر است (رحمدل، ۲۰۱۵).

داده‌کاوی، یک شیوه خودکار استخراج روابط ناشناخته و الگوهای پنهان در حجم زیادی از داده‌ها است که شامل سه تکنیک طبقه‌بندی، خوشه‌بندی و الگوکاوی می‌شود. طبقه‌بندی، فرآیند ساخت مدلی است که روی یکسری از داده‌هایی که برچسب آن از پیش تعیین‌شده آموزش می‌بیند، سپس بر اساس آن برچسب داده‌های ناشناخته را پیش‌گویی می‌کند.

درواقع طبقه‌بندی فرآیندی دومرحله‌ای است. در گام اول، یک مدل بر اساس مجموعه داده‌های آموزشی موجود در پایگاه داده ساخته می‌شود. مجموعه داده‌های آموزشی از نمونه‌ها و مثالهایی تشکیل شده‌اند که هر کدام شامل مجموعه‌ای از ویژگیها هستند. هر نمونه در مجموعه آموزش یک برچسب کلاس معلوم دارد. سیستم بر اساس این مجموعه آموزشی به خود آموزش می‌دهد یا به عبارتی پارامترهای طبقه‌بندی را برای خود مهیا می‌کند. گام بعدی پس از آموزش، پیش‌بینی یعنی تعیین برچسب کلاس نمونه‌های جدید است. درواقع

مدل ساخته شده می‌تواند برای پیشگویی برچسبهای کلاس برای داده‌های جدید مورداستفاده قرار گیرد.

در این پژوهش از طبقه‌بندهای ماشین بردار پشتیبان (SVM)، بیزین ساده (NB)، رگرسیون لجستیک (ME)، درخت تصمیم (J٤٨) و k نزدیک‌ترین همسایه (IBk) به منظور پیش‌بینی تکرار جرم استفاده شده است که در ادامه به طور مختصر شرح داده شده‌اند. شرح مفصل این الگوریتمها در پژوهشهای گذشته آورده شده است (گناناپریا، سوگانیا، دوی و کومار ، ۲۰۱۰).

ماشین بردار پشتیبان به یکی از رایج‌ترین الگوریتمهای پیش‌بینی در داده‌کاوی تبدیل شده است که در سالهای اخیر به کارایی بهتری در مقایسه با دیگر روشهای طبقه‌بندی دست یافته است. اساس کاری ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است که برای داده‌های با ابعاد بالا پیش‌بینیهای موفقیت‌آمیز دارد و صحت آن نسبت به سایر طبقه‌بندهای شناخته شده بالاتر است.

الگوریتم بیزین ساده بر پایه قضیه بیز برای مدل‌سازی پیش‌گویانه ارائه شده است. طبقه‌بندی بیز سریع‌تر از رگرسیون لجستیک همگرا می‌شود و برای طبقه‌بندی دودویی و چندگانه نتایج دقیقی ارائه می‌دهد.

رگرسیون لجستیک، یک مدل آماری رگرسیون برای متغیرهای وابسته دو سویی است. منظور از دو سویی بودن، رخداد یک واقعه تصادفی در دو موقعیت ممکنه مانند تکرار جرم یا عدم تکرار جرم است. این مدل را می‌توان به عنوان مدل خطی تعمیم‌یافته‌ای که از تابع لوجیت به عنوان تابع پیوند استفاده می‌کند و خطایش از توزیع چندجمله‌ای پیروی می‌کند، به حساب آورد.

درخت تصمیم از کاربردی‌ترین و محبوب‌ترین روشهای طبقه‌بندی محسوب می‌شود. این درخت یک روش گرافیکی قابل‌درک است که برای تصمیمات با هزینه بالا و خطرات

زیاد مورداستفاده قرار می‌گیرد. درخت تصمیم به کمک مجموعه‌ای از قوانین به پیش‌بینی مقادیر متغیر هدف می‌پردازد. در این مطالعه از الگوریتم درخت تصمیم J۴۸ در وکا که همان پیاده‌سازی درخت تصمیم معروف به C۴.۵ است استفاده می‌شود.

الگوریتم K نزدیک‌ترین همسایه که تحت عنوان جستجوی مجاورت نیز شناخته می‌شود یک مسئله بهینه‌سازی است که به پیدا کردن نزدیک‌ترین نقطه‌ها در فضاهای متریک می‌پردازد.

الگوها روابط پنهان و ناشناخته درون داده‌ها را توصیف می‌کنند. الگوکاوی یکی از روشهای شناخته شده در استخراج الگوها کاوش قواعد انجمنی[1] است که کمک می‌کند تا بتوان به‌صورت خودکار حجم زیادی از داده‌ها را تحلیل کرد و الگوهای پر رخداد این داده‌ها را استخراج کرد. قواعد انجمنی ماهیتی احتمالی دارند و به شکل اگر و آنگاه و به همراه دو معیار پشتیبان و اطمینان تعریف می‌شوند. این دو معیار به ترتیب مکرر بودن و اطمینان از قواعد مکشوفه را نشان می‌دهند. اگر قاعده‌ای، حداقل پشتیبانی را داشته باشند، «مکرر» خوانده می‌شوند. «قواعد قوی» قواعدی هستند که به طور توأمان دارای مقدار پشتیبان و اطمینان بیشتر از آستانه باشند. با استفاده از این مفاهیم کاوش قواعد انجمنی در دو گام خلاصه می‌شود: پیدا کردن مجموعه‌های مکرر و استخراج قواعد قوی. در پایان قواعد قوی به عنوان الگوهای پر رخداد مجموعه داده‌ای در نظر گرفته می‌شوند. یکی از معروف‌ترین الگوریتمهای کاوش قواعد انجمنی، اپریوری[2] نام دارد.

در سالهای اخیر، با رشد فناوری اطلاعات و گسترش سیستمهای اطلاعاتی، داده‌کاوی موردتوجه سازمانهای مختلف در سطح جهان قرار گرفته است. در این راستا تحقیقاتی نیز برای استفاده از تکنیکهای داده‌کاوی در حوزه کشف جرائم صورت گرفته است (کاپور، سینگ و چریکوری، ۲۰۲۰؛ تانگاموتو، وادیول و پریادهارشینی، ۲۰۱۹). ازجمله کارهای

1. association rule mining
2. Apriori

انجام گرفته در سطح جهان می‌توان به خوشه‌بندی جرائمی نظیر سرقت و قتل برای شناسایی سریع‌تر جرم توسط پلیس (نث ، ۲۰۰۶) و تشخیص نوع جرم و شناسایی مجرم در هند اشاره کرد (تایل و همکاران، ۲۰۱۵). ادامه این پژوهش به معرفی مطالعات انجام گرفته در ایران می‌پردازد.

کاظمی و حسین‌پور (۲۰۰۹) معتقدند که می‌توان با به‌کارگیری الگوریتمهای داده‌کاوی روابط نامحسوس داده‌های مرتبط با جرم و بزهکاری را کشف کرده و الگوهای جرم را استخراج کرد. این الگوها به پلیس کمک می‌کند تا بتواند وقوع جرم را پیش‌بینی کرده و با آرایش نظامی نیروها در منطقه جرم و کنترل دقیق‌تر آنها، از وقوع جرائم پیشگیری کند. به همین منظور آنها در پژوهش خود به بررسی تجارب و اقدامات صورت گرفته در استفاده از تکنیک داده‌کاوی برای تحلیل جرم در سازمانهای پلیسی و قضایی پرداختند.

احمدوند و آخوندزاده (۲۰۰۹) کاربرد تکنیکهای داده‌کاوی در حوزه پلیس را در سه حوزه شناسایی جرائم، پیش‌بینی جرائم و پیشگیری از جرائم را با تحلیل مطالعات گذشته موردبررسی قرار دادند. آنها دریافتند که تکنیکهای پیش‌بینی بیش از سـایر ابزارهای داده‌کاوی در این خصوص مورداستفاده قرار گرفته است. از بین الگوریتمهای پیش‌بینی، مدلهای رگرسیون حجم بیشتری از تحقیقات را به خود اختصاص داده‌اند. نتیجه تحقیق آنها ارائه یک چارچوب کاربردی برای به‌کارگیری داده‌کاوی در مسائل مرتبط با پلیس است.

اسکندری، علیزاده و کاظمی (۲۰۱۱) با به‌کارگیری ابزارهای داده‌کاوی روی بانکهای اطلاعاتی جرائم مدلی برای شناسایی و کشف جرم ارائه کردند. با استفاده از دو روش قوانین تلازمی و روش خوشه‌بندی، الگوهای موردنیاز را در شناسایی جرم سرقت کشف کردند. به‌عنوان‌مثال، نتایج نشان داد احتمال وقوع سرقت جیب بری توسط زنان ۱/۵ برابر بیشتر از مردان است. استفاده از این مدل برای پیش‌بینی جرائم دیگر به منظور پیشگیری از وقوع جرم توصیه شد.

ابراهیمی و همکاران (۲۰۱۵) از تکنیکهای داده‌کاوی برای تحلیل و بررسی اطلاعات گردآوری شده جرائم و جامعیت بخشی به بانکهای اطلاعاتی موجود استفاده کردند. آنها به کشف روابط نامحسوس میان داده‌ها و استخراج الگوهای جرم پرداختند. همچنین با استفاده از الگوریتمهای طبقه‌بندی، مدلی برای پیش‌بینی خصیصه‌های جرائم ارتکابی در آینده ارائه کردند.

در تحقیق ذکرشده از الگوریتمهای خوشه‌بندی روی مجموعه داده‌های گردآوری‌شده برای شناسایی نوع جرم استفاده شد. ارزیابی مدلهای حاصله با توجه به ویژگی هدف نوع جرم نشان داد که مدل ایجادشده توسط الگوریتم LogitBoost دارای میانگین وزن بیشتری از مدلهای دیگر است. علاوه بر این داده‌کاوی روی اطلاعات جرائم شهر لندن نشان می‌دهد که مدل ایجادشده توسط الگوریتم Bayesnet با توجه به ویژگی هدف نوع جرم دارای میانگین وزن معیار Measure-F بیشتری از مدلهای دیگری است که با استفاده از الگوریتمهای Ran-domSubSpace و IBk ارائه شدند.

صمیری و عباس‌نژاد (۲۰۱۵) با استفاده از ابزارهای داده‌کاوی مدلی برای کمک به نیروی پلیس برای پیش‌بینی وقوع جرم و درنهایت پیشگیری از آن ارائه دادند. آنها داده‌های مرتبط با جرم و بزهکاری را مورد تحلیل قرار داده، روابط پنهان میان این داده‌ها را کشف کرده و الگوهای جرم را استخراج کرده‌اند.

ابراهیم‌زاده و زرین کمری لف (۲۰۱۵) به بررسی چگونگی عملکرد کشف الگوهای جرم توسط سیستمهای داده‌کاوی پرداختند. در این تحقیق از خوشه‌بندی برای شناسایی الگوهای جرائم و همچنین از تکنیک یادگیری نیمه نظارتی برای افزایش ضریب پیش‌بینی صحیح استفاده شده است. به علاوه برای جبران محدودیتهای موجود در تکنیکها و ابزارهای خوشه‌بندی از طرح وزن‌دهی کمک گرفته شده است. نتایج تحقیق نشان داد که تکنیک مدل‌سازی در این پژوهش می‌تواند الگوهای جرم را در بین بسیاری از جرائم شناسایی کند.

البته باید به این نکته توجه داشت که این روش نیازمند یک کاوشگر ماهر و یک تحلیل‌گر متخصص برای جرم نیاز دارد.

مانیان و همکاران (۲۰۱۶) با استفاده از تکنیکهای داده‌کاوی به تحلیل داده‌های مربوط به مجرمان دستگیرشده توسط گشتهای انتظامی تهران در بهار سال ۲۰۰۹ پرداختند. آنها با کمک الگوریتمهای درخت تصمیم و شبکه عصبی روشی برای شناسایی مجرمان واقعی ارائه دادند. دلشاد و همکاران (۲۰۱۷)، با به‌کارگیری الگوریتمهای شبکه عصبی و خوشه‌بندی، گزارش‌های متنی پلیس را مورد تحلیل قرار دادند. همچنین از شبکه‌های عصبی برای کشف پیشدستانه جرم استفاده کردند. در این تحقیق تحلیل رفتار مجرم در زمان ارتکاب جرم به عنوان کلید فرایند کشف جرم شمرده شده است.

مطالعات دیگری نیز برای پیش‌بینیهای کوتاه‌مدت جرم مبتنی بر مکان، از مدل الگوریتم جنگل تصادفی (Random Forest) استفاده کرده‌اند (مولر و پورتر، ۲۰۱۸). ویلر و استین بیک (۲۰۲۱) نیز ازاین‌روش استفاده کرده و پیش‌بینیهای بلندمدتی از جرم در مکانهای کوچک ارائه دادند. آنها در این روش با استفاده از درخت تصمیم جرم را شناسایی کردند. در مطالعه آنها پیش‌بینی جرم بر اساس نزدیکی به مراکز حمل‌ونقل عمومی صورت می‌پذیرد.

برخی دیگر، از تجزیه‌وتحلیل‌های سلسله مراتبی نزدیک‌ترین همسایه (HNNC) که به طور مکرر خوشه‌های مرتبه اول و بالاتر را شناسایی می‌کند برای شناسایی نقاط با احتمال جرم بالا و بررسی همپوشانی انواع مختلف جرم استفاده کردند. به این طریق که خوشه‌های مرتبه اول با گروه‌بندی نقاطی که در فضا نزدیک‌تر از فاصله آستانه حداقل هستند ایجاد می‌شوند. سپس خوشه‌های مرتبه اول را می‌توان بر اساس فواصل میانشان، به خوشه‌های مرتبه دوم گروه‌بندی کرد. الگوریتم همچنان به تلاش برای گروه‌بندی خوشه‌ها در کلاسترهای درجه بالاتر ادامه می‌دهد تا زمانی که شکست بخورد. نتایج آنها پیامدهای مهمی برای جرم‌شناسی جغرافیایی دارند. بااین‌حال در تکنیک HNNC پارامترهای ورودی،

ذهنی بوده و لذا نمی‌توان از این تحلیلها با اطمینان نتیجه‌گیری اساسی کرد (هابرمن، ۲۰۱۷).

اخیرا، او و ژنگ (۲۰۲۱) مطالعه‌ای در خصوص پیش‌بینی میزان جرم و جنایت بر اساس یادگیری ماشینی انجام داده‌اند. آنها به منظور پیش‌بینی میزان جرم و جنایت شهر در محله‌های مختلف از شبکه‌های عصبی استفاده کردند؛ و استفاده از ایده‌های این مقاله را برای پیش‌بینی جرائم متعدد پیشنهاد کردند.

همچنین راتکلیف و همکاران (۲۰۲۱) از الگوریتمهای یادگیری ماشین استفاده کردند. این الگوریتم ابتدا از یک ماشین تقویت‌کننده گرادیان استفاده می‌کند که پیش‌بینی کند که آیا جرمی در هر سلول شبکه رخ می‌دهد یا خیر. این فرآیند متکی بر چندین سال داده‌های جرم تاریخی است که به مجموعه داده‌های آموزشی و آزمایشی تقسیم می‌شوند. بخشی از مجموعه داده آموزشی به درخت تصمیم داده می‌شود و پیش‌بینی می‌کند که آیا جرمی در یک مکان خاص (سلول شبکه) رخ خواهد داد یا خیر. این آزمایش کاهشهایی در جرم اموال ناشی از گشتهای متمرکز بر ماشین مشخص‌شده پیدا کرد. بااین‌حال تخمین و جلوگیری از جرم و جنایت در مناطق خرد انجام شده است.

مطالعه‌ای دیگر استفاده از روش نقشه‌برداری جرم و جنایت را برای پیشگیری از جرم بررسی کرد که می‌توان از آنها برای پیشگیری از جرم با استفاده از روش نقشه‌برداری جرم و جنایت روشهای بهتری نسبت به ساختارهای موجود استفاده کرد.

ابتدا تکنیکهای Crime Mapping با روشهای مختلف شناسایی و برچسب‌گذاری کانونهای جرم در نظر گرفته می‌شوند. سپس تکنیکهای پلیسی پیش‌بینی شده با مفاهیم مختلف برای طبقه‌بندی جرم موردبررسی قرار می‌گیرد. نظریه‌های جرم‌شناسی زیربنایی نیز موردبررسی و بحث قرار می‌گیرند. درنهایت، رویکردهای ممکن برای مقاوم‌سازی این تکنیکها در شهرهای هوشمند در نظر گرفته می‌شوند تا راه‌حلی جامع برای مشکل حل جرم ارائه کنند (ویلر و استین بیک، ۲۰۲۱).

هرچند مطالعات زیادی در خصوص استفاده از داده‌کاوی در زمینه‌های مختلف توسط محققان صورت گرفته است (نوفرستی، شمشادی نژاد و حیدری، ۲۰۱۸)؛ اما مطالعات موجود در زمینه داده‌کاوی جرائم در کشور ما به حوزه مواد مخدر تعمیم نیافته است. لذا این تحقیق در نظر دارد به طور خاص به شناسایی افراد مستعد به قاچاق مواد مخدر در استان سیستان و بلوچستان با استفاده از تکنیکهای داده‌کاوی بپردازد.

روش

با توجه به متغیرهای موجود در پرونده‌های مجرمان مدل مفهومی پژوهش مطابق شکل (۱) رسم شده است. همان‌طور که مشاهده می‌شود ۱۶ متغیر (ویژگی) تحقیق انتخاب شده است که در چهار رده متغیرهای شخصی، متغیرهای مرتبط با جرم، متغیرهای مرتبط با مجرم و متغیرهای زمانی دسته‌بندی شده‌اند. پایایی پرسشنامه با محاسبه آلفای کرونباخ (α) که یکی از پرکاربردترین ابزارهای آماری برای شناسایی میزان پایایی یک متغیر است محاسبه شد. مقدار آن برای کل آزمون، بالاتر از ۰/۷ محاسبه شد؛ که حاکی از پایایی مناسب است. علاوه بر این، برای تعیین روایی پرسشنامه از نظرات سه نفر متخصص در حوزه جرائم مواد مخدر استفاده شد. توصیف متغیرها و نحوه انتخاب آنها در ادامه به تفصیل آمده است. مطابق شکل (۱)، بر اساس متغیرهای تحقیق به پیش‌بینی و طبقه‌بندی متغیر هدف (متغیر دو سویی با دو مقدار تکرار جرم و عدم تکرار جرم) پرداخته می‌شود. همچنین بر اساس مقادیر متغیرهای تحقیق الگوهای پنهان استخراج شده و به عنوان خروجی فرآیند کشف دانش مورداستفاده قرار می‌گیرد.

شکل (۱) مدل مفهومی

برای انجام این تحقیق از متدولوژی استاندارد CRISP-DM[1] و الگوریتمهای دادهکاوی ماشین بردار پشتیبان[2]، بیزین ساده[3]، رگرسیون لجستیک[4]، درخت تصمیم[5] و k نزدیکترین همسایه[6] و برای استخراج الگوهای جرم از الگوریتم اپریوری[7] استفاده شده است. برای تحلیل دادهها نیز نرمافزار دادهکاوی وکا بکار گرفته شده است.

چرخه حیات یک پروژه دادهکاوی در متدولوژی CRISP_DM از شش مرحله تشکیل شده است. توالی مراحل مستقیم نیست و حرکت به عقب و جلو بین مراحل مختلف همیشه نیاز است. خروجی هر مرحله مشخص میکند که بعد از آن باید چه مرحلهای اجرا شود (نوفرستی و شمس فرد، ۲۰۱۵). بردارها وابستگیهای مهم بین مراحل را مشخص میکند. این چرخه در شکل ۲ نمایش داده شده است. در ادامه خلاصه هر مرحله پروژه حاضر بیان میشود.

1. cross industry standard process for data mining
2. support vector machine
3. Naïve Bayes
4. logistic regression
5. decision tree
6. k Nearest Neighbors
7. Apriori

شکل (۲) مراحل انجام پژوهش

شناخت مسئله: در این مرحله به شناخت سیستم و بیان اهداف مسئله مورد نظر پرداخته می‌شود. استان سیستان و بلوچستان به لحاظ وضعیت جغرافیایی، اقتصادی و اجتماعی بیشتر در معرض خطر آلودگی به مواد مخدر به دو شکل اعتیاد و قاچاق قرار دارد. آمار بالای قاچاق مواد مخدر در استان سیستان و بلوچستان و صدمات جبران‌ناپذیر آن بر افراد و خانواده‌های درگیر، شناسایی افراد مستعد به قاچاق مواد مخدر در این استان را حائز اهمیت می‌کند (رحیمی و کاویان، ۲۰۱۵). هدف تحقیق حاضر، ارائه مدل پیش‌بینی کننده‌ای برای شناسایی افراد مستعد به قاچاق مواد مخدر و نیز استخراج الگوهای جرائم است.

جمع‌آوری و درک داده: به منظور استفاده از ابزارها و الگوریتمهای داده‌کاوی نیاز به یک مجموعه داده‌ای قابل تفسیر توسط کامپیوتر است. در این راستا و با توجه به محدودیتهای امنیتی در این حوزه، با همکاری و هماهنگی شورای مواد مخدر استانداری سیستان و بلوچستان، مجموعه داده‌های واقعی با استفاده از روش نمونه‌گیری در دسترس از پرونده‌های ۵۳۹ مجرم حوزه مواد مخدر در استان سیستان و بلوچستان که در طی سالهای ۲۰۱۳ الی ۲۰۲۰ مرتکب جرم قاچاق مواد مخدر شده و اطلاعات مربوط به آنان ثبت شده است،

جمع‌آوری شد. به منظور حفظ حریم خصوصی افراد، اطلاعات هویتی مجرمان شامل نام، نام خانوادگی، نام پدر، شماره پرونده و شماره شناسنامه از پایگاه داده حذف شد.

آماده‌سازی داده‌ها: این مرحله شامل چهار گام استخراج ویژگی، پاک‌سازی داده‌ها، یکپارچه‌سازی داده‌ها و تبدیل داده‌ها می‌شود. در گام اول با مشورت کارشناسان حوزه مواد مخدر و بر اساس مطالعات پیشین، ١٦ خصیصه جنسیت مجرم، رده سنی مجرم، وضعیت تأهل مجرم، سطح تحصیلات مجرم، محل تولد مجرم، نوع جرم، شریک داشتن مجرم، نوع ارتباط مجرم با همدستان، شغل مجرم، وضعیت مالی مجرم، فصل وقوع جرم، وضعیت سوءپیشینه مجرم، دلیل ارتکاب جرم، تعداد دفعات تکرار جرم، نوع و مقدار ماده مخدر برای پیش‌بینی و شناسایی افراد مستعد به قاچاق مواد مخدر انتخاب شد. بر اساس پرونده مجرمان، مواد مخدر قاچاق شده توسط این افراد شامل تریاک، شیره، حشیش، مرفین و شیشه بوده است. جدول (۱) ویژگیهای تحقیق را به همراه نوع آنها نشان می‌دهد. همچنین مجموعه مقادیر یا بازه مقادیر مشاهده شده برای هر ویژگی در داده‌های موردبررسی نشان داده شده است.

در گام دوم داده‌های ناقص و دارای خطا حذف شدند. در این راستا، پس از بررسی ۵۳۹ پرونده و جمع‌آوری اطلاعات تکمیلی از طریق پرسشنامه، مواردی که دارای داده ناقص یا متناقض بودند کنار گذاشته شدند. در پایان اطلاعات ٤٦٧ نفر که پرسشنامه را به طور کامل پاسخ داده بودند و مؤلفه‌های لازم برای استفاده در این تحقیق را داشتند، انتخاب شد.

در گام سوم داده‌های جمع‌آوری شده از منابع مختلف به شکل واحد در آمدند و در گام چهارم داده‌ها به فرمت مناسب برای پردازش توسط ماشین تبدیل شدند. به‌عنوان‌مثال، به دلیل پراکندگی مقادیر ویژگی سن، سن افراد به چهار بازه زیر ۱۸ سال، ۱۸ تا ۲۲ سال، ۲۳ تا ۳۰ سال، ۳۱ تا ٤۰ سال و بالای ٤۰ سال تقسیم شده است. به طور مشابه ویژگی مقدار ماده مخدر همراه مجرم به تعدادی بازه تقسیم شده است. دسته‌بندیهای مذکور در جدول (۱) نشان داده شده‌اند.

جدول (۱) ویژگیهای مورداستفاده و نوع آنها

ردیف	ویژگی	نوع	محدوده مقادیر
۱	جنسیت مجرم	اسمی	{زن، مرد}
۲	رده سنی مجرم	اسمی	{کمتر از ۱۸ سال، ۱۸ تا ۲۲ سال، ۲۳ تا ۳۰ سال، ۳۱ تا ٤٠ سال و بالای ٤٠ سال}
۳	وضعیت تأهل مجرم	اسمی	{مجرد، متأهل}
٤	سطح تحصیلات مجرم	اسمی	{بی‌سواد، ابتدایی، راهنمایی، دبیرستان، دیپلم و تحصیلات دانشگاهی}
٥	محل تولد مجرم	اسمی	{خاش، زاهدان، زابل، ایرانشهر، مشهد، کرمان، سراوان و سایر}
٦	نوع جرم	اسمی	{فروش مواد مخدر، فروش و مصرف مواد مخدر}
۷	شریک	اسمی	{دارد، ندارد}
۸	نوع ارتباط مجرم با همدستان	اسمی	{صاحب مال، همدست، کرایه‌ای}
۹	شغل مجرم	اسمی	{بیکار، شغل آزاد، کارگر، کارمند}
۱۰	وضعیت مالی مجرم	اسمی	{ضعیف، متوسط، خوب}
۱۱	فصل وقوع جرم	اسمی	{بهار، تابستان، پاییز، زمستان}
۱۲	سوءپیشینه	اسمی	{دارد، ندارد}
۱۳	دلیل ارتکاب جرم	اسمی	{خانواده، فقر، بیکاری، کسب درآمد، دوست ناباب}
١٤	تعداد دفعات تکرار جرم	عددی	[۱ ... ۱۰]
۱٥	نوع ماده مخدر	اسمی	{تریاک و شیره، شیشه، حشیش و مرفین}
١٦	مقدار ماده مخدر	اسمی	{کمتر از ۱۰۰ گرم، ۱۰۰ تا ۲۰۰، ۲۰۰ تا ٤٠٠، ٤٠٠ تا ٦٠٠، ٦٠٠ تا ۸۰۰، ۸۰۰ تا ۱۰۰۰، ۱۰۰۰ تا ۲۰۰۰، بیشتر از ۲۰۰۰}

مدلسازی: در این مرحله، تکنیکهای مختلف مدلسازی انتخاب و بکار گرفته می‌شوند. در این تحقیق در ابتدا با استفاده از تکنیک طبقه‌بندی به تولید مدل و الگوی بهینه‌ای برای پیش‌بینی افراد مستعد به قاچاق مواد مخدر پرداخته شده است. سپس با استفاده از الگوکاوی الگوهای جرم استخراج می‌شوند.

یک مجموعه آموزش ایجاد شده است که در آن هر نمونه دارای برچسب کلاس تکرار یا عدم تکرار جرم است. سیستم بر اساس این مجموعه آموزشی به خود آموزش می‌دهد و یک مدل پیشگویانه می‌سازد. از این مدل برای پیش‌بینی تکرار جرم توسط فرد استفاده شده است. به بیانی دیگر با دادن ۱۶ ویژگی ذکر شده در بخش آماده‌سازی داده‌ها یا بخشی از این ویژگیها، سامانه طراحی‌شده پیش‌بینی می‌کند که فرد مذکور مجدداً مرتکب جرم می‌شود یا خیر. درواقع سامانه طراحی‌شده قادر به شناسایی افراد مستعد قاچاق مواد مخدر است.

در این پژوهش از طبقه‌بندهای ماشین بردار پشتیبان، بیزین ساده، رگرسیون لجستیک، درخت تصمیم و k نزدیک‌ترین همسایه به منظور پیش‌بینی تکرار جرم برای نمونه‌های جدید (متهمانی که اخیراً دستگیر شده‌اند) استفاده شده است. علت انتخاب طبقه‌بندهای مذکور، رایج بودن استفاده از آنها در کاربردهای مختلف و موفقیت‌آمیز بودن نتایج حاصل بوده است. دقت طبقه‌بندهای مذکور در شناسایی افراد در معرض خطر در بخش تجزیه‌وتحلیل یافته‌ها ارائه شده است. برای پیاده‌سازی الگوریتمهای مذکور از نرم‌افزار وکا استفاده شده است.

در گام دوم از مدلسازی به استخراج الگوهای جرم با استفاده از الگوریتم اپریوری پرداخته می‌شود.

ارزیابی مدل: برای ساخت مجموعه آموزش و مجموعه تست از تکنیک رایج ارزیابی متقابل با ۱۰ حلقه[1] استفاده شد. در این تکنیک مجموعه داده‌ها به ۱۰ قسمت مساوی تقسیم می‌شود که در هر تکرار از الگوریتم یک بخش به عنوان تست و ۹ بخش دیگر به عنوان

1. 10-Fold Cross Validation

آموزش انتخاب می‌شوند. نهایتاً میانگین ۱۰ بار تکرار الگوریتم به عنوان نتیجه نهایی انتخاب می‌شود.

به منظور مقایسه مدلهای ساخته شده از معیار صحت استفاده شده است که به‌صورت زیر تعریف می‌شود:

(۱)صحت $$\frac{TP+TN}{TP+FP+TN+FN}$$

متغیر TP یا «مثبت صحیح» به تعداد رکوردهای مثبت (افراد مستعد به قاچاق مواد مخدر) از مجموعه تست اشاره دارد که توسط الگوریتم داده‌کاوی به‌درستی دسته‌بندی شده‌اند یعنی برچسب مثبت (تکرار جرم) گرفته‌اند. به طور مشابه متغیر TN یا «منفی صحیح» به تعداد رکوردهای منفی (افرادی که تکرار جرم نداشته‌اند) از مجموعه تست اشاره دارد که توسط الگوریتم داده‌کاوی به‌درستی دسته‌بندی شده‌اند یعنی برچسب منفی (عدم تکرار جرم) گرفته‌اند. متغیر FP یا «مثبت کاذب» نیز تعداد رکوردهای منفی از مجموعه تست که اشتباهاً به عنوان مثبت دسته‌بندی شده‌اند را نشان می‌دهد. به بیانی دیگر، FP تعداد افرادی است که در واقعیت تکرار جرم نداشته‌اند اما توسط مدل ساخته شده به‌اشتباه برچسب تکرار جرم خورده‌اند. FN یا «منفی کاذب» تعداد رکوردهای مثبت مجموعه تست که به‌اشتباه برچسب منفی خورده‌اند را نشان می‌دهد. درواقع FN تعداد افرادی است که تکرار جرم داشته‌اند اما مدل پیش‌بینی نتوانسته است آنها را تشخیص دهد.

علاوه بر صحت، شاخصهای حساسیت (فراخوانی)، ارزش اخباری مثبت (دقت) و معیار F نیز برای مقایسه مدلها مورداستفاده قرار گرفته‌اند که در ادامه فرمول محاسبه هر کدام ارائه شده است.

(۲) حساسیت $$\frac{TP}{TP+FN}$$

(۳) ارزش اخباری مثبت $\frac{TP}{TP+FP}$

(٤) فراخوانی* دقت*۲ / فراخوانی* دقت

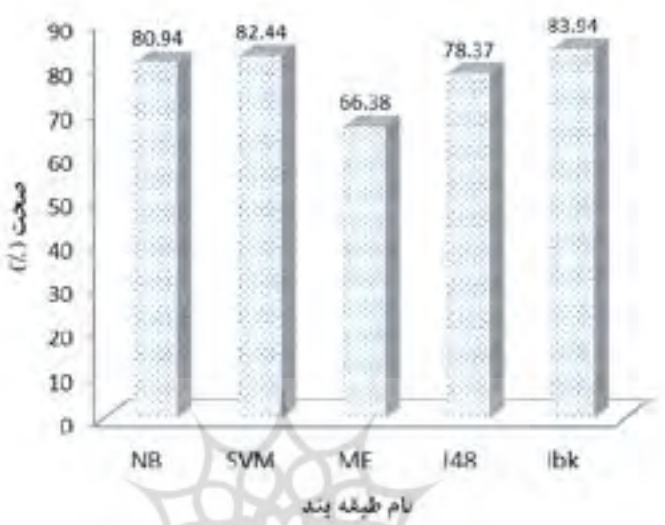
توسعه: با به‌کارگیری طبقه‌بند IBk سامانه‌ای برای پیش‌بینی افراد مستعد به قاچاق موارد مخدر طراحی شد و همچنین برای تأیید قواعد و الگوهای استخراج شده، از کارشناسان حوزه مواد مخدر کمک گرفته شده است. بدین ترتیب الگوهای تأییدشده توسط کارشناسان برای تحلیل دلایل اصلی وقوع جرم و ویژگیهای مجرمان به مسئولان مربوطه ارائه شد.

یافته‌ها

به منظور پیش‌بینی تکرار جرم با استفاده از الگوریتمهای طبقه‌بندی، در ابتدا با روش ارزیابی متقابل با ده حلقه، مجموعه داده‌ای به دو بخش آموزش و تست تقسیم شده است. طبقه‌بند یادگیری ماشین روی مجموعه آموزش، یادگیری انجام می‌دهد. سپس بر اساس مدل یادگرفته شده، به پیش‌بینی تکرار جرم در مجموعه تست می‌پردازد. بر اساس درصدی از نمونه‌های مجموعه تست که پیش‌بینی تکرار جرم در آنها به‌درستی تخمین زده شده است، صحت طبقه‌بند محاسبه می‌شود.

در شکل (۳) صحت طبقه‌بندهای مذکور در پیش‌بینی تکرار جرم توسط یک شخص نمایش داده شده است. همان‌طور که مشاهده می‌شود، طبقه‌بند IBk به صحت بالاتری در مقایسه با سایر طبقه‌بندها دست یافته است. طبقه‌بند IBk با صحت حدود ۸٤ درصد قادر به پیش‌بینی تکرار جرم است.

شکل (۳) مقایسه صحت طبقه‌بندها در پیش‌بینی تکرار جرم



جدول (۲) کارایی طبقه‌بندها در شناسایی افراد مستعد به قاچاق مواد مخدر را بر اساس شاخصهای مختلف شامل صحت، حساسیت، ارزش اخباری مثبت و معیار F مقایسه می‌کند. همان‌طور که مشاهده می‌شود طبقه‌بند IBk بر اساس همه شاخصهای مذکور بر سایر طبقه‌بندها برتری داشته است.

جدول (۲) مقایسه طبقه‌بندها در پیش‌بینی تکرار جرم بر اساس شاخصهای مختلف

طبقه‌بند	صحت	حساسیت (فراخوانی)	ارزش اخباری مثبت (دقت)	معیار F
NB	۸۰/۹٤	۸۰/۹	۸۰/٦	۸۰/۸
SVM	۸۲/٤٤	۸۲/٤	۸۱/۲	۸۱/٦
ME	٦٦/۳۸	٦٦/٤	۷٤/۸	٦۹/۱
J48	۷۸/۳۷	۷۸/٤	٦۲	٦۹/۲
IBk	۸۳/۹٤	۸۳/۹	۸۳/۲	۸۳/٤

در جدول (۳)، برخی از الگوهای جرم استخراج شده توسط الگوریتم الگوکاوی نمایش داده شده است. به‌عنوان‌مثال الگوی شماره (۱) بیان می‌کند که افرادی که شریک جرم ندارند و وضعیت مالی آنها ضعیف است و نوع ماده مخدر همراه آنها تریاک بوده است عمدتاً متأهلند. برای هر الگو رخداد و درصد اطمینان نیز نشان داده شده است. رخداد یا پشتیبان یعنی تعداد دفعاتی که الگو در پایگاه داده مشاهده شده است و اطمینان یعنی درصد مواردی که الگو درست بوده است. به طور مشابه، الگوی شماره (۲) مبین این است که مجردهایی که سوءپیشینه داشته‌اند با احتمال ۹۶ درصد بیکارند.

جدول (۳) الگوهای جرم استخراج شده

شماره	الگو	رخداد	اطمینان
۱	شریک=ندارد، وضعیت مالی=ضعیف، نوع ماده مخدر=تریاک ←وضعیت تأهل=متأهل	۹۵	٪۹۹
۲	وضعیت تأهل=مجرد، سوءپیشینه=دارد← شغل=بیکار	۹۱	٪۹۶
۳	محل تولد=زاهدان، وضعیت تأهل=متأهل، وضعیت مالی=ضعیف ← تکرار جرم=دارد	۷۵	٪۹۴
٤	رده سنی=٤۰-۳۱، وضعیت تأهل=متأهل ← تکرار جرم=دارد	۵۸	٪۹۵
۵	دلیل ارتکاب جرم=کسب درآمد، وضعیت تأهل=متأهل، نوع ماده مخدر=تریاک ← تکرار جرم=دارد	۷٤	٪۹۱
٦	محل تولد=زابل، نوع ماده مخدر=تریاک ← رده سنی=۲۳ تا ۳۰ سال	۵۹	٪۸٤
۷	رده سنی=٤۰-۳۱ سال، تکرار جرم=دارد ← وضعیت مالی=ضعیف	۵۸	٪۸٤
۸	شغل=آزاد، وضعیت مالی=خوب ← نوع مواد=تریاک	۵٦	٪۹۲

با تحلیل الگوهای استخراج شده توسط کارشناسان حوزه مواد مخدر می‌توان دلایل اصلی ارتکاب جرم و افراد مستعد تکرار جرم را شناسایی و راهکارهایی برای کاهش جرائم ارائه داد. همچنین با تحلیل روانشناسانه الگوها توسط مددکارهای اجتماعی و

سایر متخصصان مربوطه، می‌توان تا حد زیادی از ارتکاب به جرائم مشابه پیشگیری کرد. به‌عنوان‌مثال الگوهای شماره (۳) تا (۵) بیان می‌کنند که عمدتاً افراد متأهل متولد زاهدان که وضعیت مالی بدی دارند، متأهلین رده سنی ۳۱ تا ٤۰ سال و نیز متأهلینی که با هدف کسب درآمد تریاک قاچاق می‌کنند، مجدداً مرتکب جرم می‌شوند. بنابراین به منظور کاهش جرائم مرتبط با قاچاق مواد مخدر، بهبود وضعیت معیشتی مردم استان، ایجاد فرصتهای شغلی به ویژه برای متأهلین و نیز رسیدگی به مشکلات و دغدغه‌های افراد در رده سنی ۳۱ تا ٤۰ سال باید در اولویت قرار گیرد.

الگوی شماره (٦) نشان می‌دهد که ۸٤ درصد افرادی که تریاک قاچاق کرده‌اند و متولد زابل هستند در رده سنی ۲۳ تا ۳۰ سال قرار دارند.

الگوی شماره (۷) بیان می‌کند که در رده سنی ۳۱ تا ٤۰ سال، افرادی که چندین بار مرتکب قاچاق مواد مخدر شده‌اند، وضعیت مالی بدی داشته‌اند. این الگو مبین این است که یکی از دلایل قاچاق مواد مخدر در رده سنی ۳۱ تا ٤۰ سال که عمدتاً متأهل هستند و بار اقتصادی یک خانواده را به دوش می‌کشند، وضعیت بد اقتصادی است. بنابراین رسیدگی به وضعیت معیشتی مردم و ایجاد فرصتهای شغلی توسط مسئولین می‌تواند راهگشا باشد.

در مقابل، الگوی شماره (۸) نشان می‌دهد افراد دارای مشاغل آزاد که وضعیت مالی خوبی داشته‌اند عمدتاً تریاک حمل می‌کنند. این الگو مبین این است که حتی برخی از افرادی که وضعیت مالی مناسبی دارند نیز مرتکب جرائم مرتبط با مواد مخدر می‌شوند. علت اصلی این امر می‌تواند ناآگاهی این افراد از عوارض سوء مواد مخدر، تفریح و یا همنشینی با دوستان ناباب باشد. در این راستا راهکارهایی نظیر آگاه‌سازی و آموزش مردم با ساخت برنامه‌های تلویزیونی، همچنین پخش بروشورهای هشداردهنده به ویژه در محلهای تجمع عمومی در روستاها مانند مساجد و محل برگزاری نماز جمعه و گسترش تفریحات سالم در سطح استان به ویژه در رده سنی جوانان می‌تواند مؤثر باشد.

بحث

با توجه به اهمیت مسئله قاچاق مواد مخدر در استان سیستان و بلوچستان، ارائه راهکارهایی برای شناسایی الگوهای بیانگر دلایل وقوع جرائم و نیز شناسایی افراد در معرض خطر از ضروریات است. ازاین‌رو در این مقاله، با تحلیل داده‌های مربوط به مجرمان مواد مخدر با استفاده از الگوریتمهای طبقه‌بندی، سامانه‌ای برای تشخیص افراد مستعد به قاچاق مواد مخدر (یعنی افرادی که مستعد تکرار جرم) هستند طراحی شد. نتایج ارزیابیهای انجام گرفته نشان می‌دهد طبقه‌بند IBk قادر است با دقت ۸٤ درصد افراد در معرض خطر را شناسایی کند. همچنین با به‌کارگیری الگوریتم الگوکاوی، الگوهای جرائم استخراج شده است. این الگوها می‌توانند کمک شایانی به مسئولان مربوطه برای تشخیص افراد در معرض خطر و نیز شناسایی دلایل ارتکاب جرم باشند.

ازجمله محدودیتهایی که طی انجام تحقیق حاضر وجود داشت، مشکلات فراوان در مسیر جمع‌آوری داده بود. ثبت سیستمی اطلاعات مجرمان بجای استفاده از پرونده می‌تواند تا حد زیادی از مشکلات جمع‌آوری و آماده‌سازی داده و نیز خطای انسانی ناشی از ثبت دستی اطلاعات بکاهد و باعث صرفه‌جویی در زمان شود. به علاوه با در اختیار داشتن اطلاعات بیشتر مانند یک بانک اطلاعات جامع مکانی که علاوه بر مشخصات جمعیت‌شناختی مجرمان و خصیصه‌های جرم، محل دقیق وقوع جرم و دستگیری مجرم را نیز در خود جای دهد، می‌توان الگوهای بهتر و دقیق‌تر ازجمله الگوهای مکانی جرم را نیز استخراج کرد.

در تحقیقات آتی می‌توان دیگر الگوریتمهای داده‌کاوی را نیز برای پیش‌بینی افراد مستعد به قاچاق مواد مخدر مورداستفاده قرار دارد. به علاوه می‌توان به روشی مشابه به تحلیل و پیش‌بینی سایر جرائم پرداخت.

ملاحظات اخلاقی

مشارکت نویسندگان

همه نویسندگان در تهیه مقاله مشارکت داشته‌اند.

منابع مالی

برای انتشار این مقاله حمایت مالی دریافت نشده است.

تعارض منافع

مقاله حاضر تعارض منافع با سایر مقالات نویسندگان ندارد.

پیروی از اصول اخلاقی پژوهش

در این مقاله همه حقوق مرتبط با اخلاق پژوهش رعایت شده است.

تشکر و قدردانی

بدین‌وسیله گروه تحقیق از همکاری شورای هماهنگی مبارزه با مواد مخدر استانداری استان سیستان و بلوچستان تشکر و قدردانی می‌کند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

- Ahmadvand, A. & Akhondzadeh, M.A. (2010). Framework for the application of data mining techniques in crime modeling, ***Police Human Development Bi-Quarterly***, 17(30), 11-22. (In Persian).

- Bakhtiari Shahri, A., & Eftekhari, N. (2017). ***Detection of people prone to drug trafficking in Sistan and Baluchestan province using data mining techniques***. Fourth National Conference on Information Technology, Computer and Telecommunications, Torbat-e Heydarieh. (In Persian).

- Delshad, M., Tudeh Zaeem, B., & Rastegar, A. (2017). ***Computer analysis of crime using artificial intelligence and pre-crime detection data mining methods***. National Conference on Vision 1420 and Technological Advances in Electrical Engineering, Computer and Information Technology, Shiraz, Iran New Education Development Center. (In Persian).

- Ebrahimi, R.V., & Bafandrh, H. (2017). ***Components of Temperament and Emotional Intelligence in People Addicted to Drugs and Normal***. Paper presented at the Contemporary Psychology.

- Ebrahimi, M., Mirroshandel, S. A., & Aghaii, J. A. (2015). Comprehensive part of the crime database in order to predict and identify crimes using data mining techniques. ***Quarterly Journal of Electronics Industry***, 4 (6), . (In Persian).

- Ebrahimzadeh, S., & Zarrin Kamri Lef, M. (2015), ***Investigation of how, function, crime pattern detection by data mining systems***. 1404 National Conference on Vision and Technological Achievements of Engineering Sciences. (In Persian).

- Eskandari, H. Alizadeh, S., & Kazemi, P. (2011). Application of data mining in identifying and discovering hidden patterns of theft. ***Journal of Law Enforcement and Security***, 4 (4), 35-56. (In Persian).

- Gholami, H., & Barzegar, M. (2018). ***The Function of Methods for Prediction of Recidivism in Grant of Parole***.

- Gnanapriya, S., Suganya, R., Devi, G., & Kumar, M. (2010). Data Mining Concepts and Techniques. ***Data Mining and Knowledge Engineering***, 2(9), 256-263.

- Haberman, C.P. (2017). Overlapping hot spots? Examination of the spatial heterogeneity of hot spots of different crime types. ***Criminology & Public Policy***, 16(2), 633-660.

- He, J., & Zheng, H. (2021). Prediction of crime rate in urban neighborhoods based on machine learning. ***Engineering Applications of Artificial Intelligence,*** 106,104460.

- Kapoor, P., Singh, P.K., & Cherukuri, A.K. (2020). Crime Data Set Analysis Using Formal Concept Analysis (FCA): A Survey Advances in Data Sciences, ***Security and Applications*** (pp. 15-31), Springer.

- Kazemi, P., & Hosseinpour, J. (2009). ***Application of data mining in police and judicial organizations in order to model crime and detect crimes.*** The Second International Conference on Electronic City, Jihad University Information and Communication Technology Research Institute, Tehran. (In Persian).

- Manian, A., Jamalo, M., & Biddle, M. (2016). Designing a proposed data mining model to identify criminals. ***Social Order Quarterly,*** 8 (3), 109-128. (In Persian).

- Mohler, G., & Porter, M.D. (2018). Rotational grid, PAI-maximizing crime forecasts. Statistical Analysis and Data Mining: ***The ASA Data Science Journal***, 11(5), 227-236.

- Nath, S.V. (2006) ***of Conference. Crime Pattern Detection Using Data Mining***. Paper presented at the International Conference on Web Intelligence and Intelligent Agent Technology Workshops.

- Noferesti, S., Shamshadinejad, N., & Haidari, F (2018). Use of data mining techniques for differential diagnosis of iron deficiency and beta thalassemia minor, ***health informatics and biomedicine***, 435-446. (In Persian).

- Noferesti, S., & Shamsfard, M. (2015). Resource construction and evaluation for indirect opinion mining of drug reviews. ***PloS one,*** 10(5): e0124993.

- Rahimi Movagher, A., & Kaviyan, M. (2015). Drug Abuse Prevention Guide for Media Staff. United Nations Office Drugs and Crime.

- Rahmdel, M. (2015). ***Iranian Criminal Policy Vis -a- Vis Drug Offenes (3 ed.): The Organization for Researching and Composing University Textbooks in the Humanities(SAMT).***

- Ratcliffe, J.H. et al. (2021). The Philadelphia predictive policing experiment. ***Journal of Experimental Criminology***, 17(1), 15-41.

- Samiri, Kh., & Abbasnejad, H. (2015). ***Combining data sources, methods of discovery and research of crime data extraction***. The First International Congress of Iranian Law. (In Persian).

- Tayal, D.K., Jain, A., Arora, S., Agarwal, S., Gupta, T., & Tyagi, N. (2015). Crime detection and criminal identification in India using data mining techniques. ***AI & society***, 30(1), 117-127.

- Thangamuthu, M.A., Vadivel, M.G., & Priyadharshini, M.A. (2019). Detecting Criminal Method using Data Mining. ***INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY***, 6(3).

- Wheeler, A.P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. ***Journal of Quantitative Criminology,*** 37(2), 445-480

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی