

DOI: <https://doi.org/10.34069/AI/2022.56.08.27>

How to Cite:

Tymchyshyn, A., Semeniaka, A., Bondar, S., Akhtyrskaya, N., & Kostiuchenko, O. (2022). The use of big data and data mining in the investigation of criminal offences. *Amazonia Investiga*, 11(56), 278-290. <https://doi.org/10.34069/AI/2022.56.08.27>

The use of big data and data mining in the investigation of criminal offences

Застосування Big Data та Data Mining у розслідуванні кримінальних правопорушень

Received: September 10, 2022

Accepted: October 07, 2022

Written by:

Andriy Tymchyshyn¹¹¹<https://orcid.org/0000-0002-9591-8273>**Anna Semeniaka**¹¹²<https://orcid.org/0000-0001-9366-8234>**Serhii Bondar**¹¹³<https://orcid.org/0000-0002-0497-4457>**Nataliia Akhtyrskaya**¹¹⁴<https://orcid.org/0000-0003-3357-7722>**Olena Kostiuchenko**¹¹⁵<https://orcid.org/0000-0002-2243-1173>

Abstract

The aim of this study was to determine the features and prospects of using Big Data and Data Mining in criminal proceedings. The research involved the methods of a systematic approach, descriptive analysis, systematic sampling, formal legal approach and forecasting. The object of using Big Data and Data Mining are various crimes, the common features of which are the seriousness and complexity of the investigation. The common tools of Big Data and Data Mining in crime investigation and crime forecasting as interrelated tasks were identified. The creation of databases is the result of the processing of data sources by Data Mining methods, each being distinguished by the specifics of use. The main risks of implementing Big Data and Data Mining are violations of human rights and freedoms. Improving the use of Big Data and Data Mining requires standardization of procedures with strict adherence to the fundamental ethical, organizational and procedural rules. The use of Big Data and Data Mining is a forensic

Анотація

Метою даного дослідження стало визначення особливостей та перспектив застосування Big Data та Data Mining в кримінальних провадженнях. У дослідженні були застосовані методи системного підходу, описового аналізу, системного відбору, формально-юридичний та прогнозування. Об'єктом застосування Big Data і Data Mining є різноманітні злочини, спільними ознаками яких є серйозність та складність розслідування. Виявлена інструментальна спільність Big Data і Data Mining в розслідуванні злочинів та прогнозуванні злочинності як взаємопов'язаних завдань. Формування баз даних є результатом опрацювання джерел методами Data Mining, кожен з яких вирізняється специфікою використання. Основні ризики впровадження Big Data і Data Mining полягають у порушеннях прав і свобод людини. Удосконалення застосування Big Data і Data Mining потребує стандартизації процедур з чітким дотриманням засадничих правил етичного, організаційного та процесуального

¹¹¹ PhD in Law Sciences, Associate Professor, Department of Law, Separate Structural Subdivision of Higher Education Institution "Open International University of Human Development Ukraine" Ivano-Frankivska Branch, Ivano-Frankivsk, Ukraine.

¹¹² Postgraduate student, Tavria National University named after V. I. Vernadskyi, Kyiv, Ukraine.

¹¹³ PhD in Law Sciences, Senior Research Fellow, Department for the organization of scientific activities and protection of intellectual property rights, National Academy of Internal Affairs, Kyiv, Ukraine.

¹¹⁴ PhD in Law Sciences, Associate Professor, Department of Criminal Process and Criminalistics, Educational and Scientific Institute of Law of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

¹¹⁵ PhD in Law Sciences, Associate Professor, Department of Criminal Process and Criminalistics, Educational and Scientific Institute of Law of Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.



innovation in the investigation of serious crimes and the creation of an evidence base for criminal justice. The prospects for widespread use of these methods involve the standardization of procedures based on ethical, organizational and procedural principles. It is appropriate to outline these procedures in framework practical recommendations, emphasizing the responsibility of officials in case of violation of the specified principles. The area of further research is the improvement of innovative technologies and legal regulation of their application.

Keywords: criminal analytics, criminal justice, criminal offenses, investigation, working with data.

Introduction

The law enforcement agencies are moving from rare partial cases of the use of modern technologies in criminal proceedings to their complex application, the development of new methods of detection and investigation of criminal offences. This is determined by a number of factors that are inherent in the vast majority of proceedings. These are the intellectualization of crimes and ways of countering their detection by criminals; significant data volumes that detectives need to process; lack of time and the dynamic investigation environment (Blahuta & Movchan, 2020). Information remains the central issue of the entire investigation process — its search, processing, consolidation and use as evidence.

Law enforcement systems generate huge volumes of information about crimes. These are demographic, socio-economic, time-space, geographic data (Butt et al., 2020). Detectives get a significant part of them from social networks, which occupy a special place among the sources of criminally significant data (Zhou et al., 2021). That is why a data management model or technique is so important for crime prevention decision-making (Hussain & Aljuboori, 2022). In this context, only the latest methods and technologies can give law enforcement agencies the opportunity to quickly and efficiently investigate and detect criminal offences.

Big Data and Data Mining occupy a special place among innovations, the use of which is determined by the specifics of the information society in which crimes are committed. In a generalized sense, Big Data is a concept that

характеру. Застосування Big Data і Data Mining є криміналістичною інновацією в розслідуванні серйозних злочинів та формування доказової бази для кримінального судочинства. Перспективи широкого застосування цих методів полягають у стандартизації процедур, що ґрунтуються на етичній, організаційній та процесуальній засадах. У рамочному виді дані процедури доцільно викласти в практичних рекомендаціях із наголошенням на відповідальності посадових осіб при порушенні зазначених засад. Орієнтири подальших розвідок лежать в сфері удосконалення інноваційних технологій та правової регламентації їх застосування.

Ключові слова: робота з даними, кримінальна аналітика, кримінальні правопорушення, розслідування, кримінальне правосуддя.

represents a huge amount of structured, unstructured and semi-structured data (Usha et al., 2020). In the era of Big Data, there is a transition to modern ways of collecting and integrating small-scale data contained in various sources (Zhao & Tang, 2017). Data Mining is a method of working with large arrays of data using computer technologies with subsequent identification of their significance, comprehensive analysis and generalization to the required informational result (Pokhriyal et al., 2020). In this context, Data Mining is a powerful tool with practical potential. Thanks to it, investigators can focus on the most important information about the crime (Hassani et al., 2016).

Aim

In view of the foregoing, the aim of this study is to consider the specifics of the use of Big Data and Data Mining in the investigation of criminal offences, as well as to determine problematic aspects in the field of human rights protection. The aim involves the following research objectives:

- identify the substance and tasks of Big Data and Data Mining as methods of investigating crimes and predicting criminal activity;
- identify criminal, procedural and human rights components of the application of these methods;
- determine prospects for the introduction of standards for the use of Big Data and Data Mining as forensic innovations in the investigation of crimes.

Literature review

The use of Big Data and Data Mining enables covering a number of well-known algorithms of intellectual analysis, which are involved in the detection and investigation of criminal offences. These include text analysis (Pramanik et al., 2017) (natural language processing (Chaudhary & Bansal, 2022), content processing through the development and application of a criminal thesaurus (Das et al., 2021), topic modelling (Zhao & Tang, 2017)); analysis of competing hypotheses during investigation (Oatley et al., 2020); studying the specifics of the connection between the crime and the territory in which it is committed (Hussain & Aljuboori, 2022), which is used to form geographic clusters (Usha et al., 2020); structural analysis of social networks (Pramanik et al., 2017), etc. Artificial intelligence is a promising Data Mining tool, which supplements the forensic capabilities of law enforcement agencies for processing information and its further analysis (Dupont et al., 2018). In particular, Data Mining by means of artificial intelligence includes Big Data processing for profiling and forecasting criminal behaviour; predicting crime rates (Oatley, 2022), etc.

The active use of Big Data and Data Mining has led to the creation of intelligent platforms for the coordination of law enforcement activities and information provision of current and planned policing (Norouzi & Ataei, 2021). One of them is the EU law enforcement agency's Secure Information Exchange Network Application (SIENA) platform. It ensures the exchange of operational and strategic information on crime between: Europol analysts and experts; EU member states; third countries with which Europol has cooperation agreements or working arrangements (Europol, 2022). Open Source Intelligence (OSINT), which is used to analyse publicly available sources of information, is one of the solutions to counter terrorist activities on the Internet (Chaudhary & Bansal, 2022).

The main focus is, however, on the technological aspects of Big Data and Data Mining. Although their application in criminal proceedings is covered, the criminal law, procedural and human rights aspects remain insufficiently studied.

In particular, it is necessary to pay attention to the types of crimes, to determine their most significant features in order to make innovations in the field of criminal justice more effective (Das et al., 2021). Specialists mainly focus on the advantages of using Big Data and Data Mining to

detect and investigate certain types of acts: fraud and other economic crimes in the business environment (Dehtiarovai & Yevdokimov, 2018), terrorist activity in social networks (Chaudhary & Bansal, 2022) etc. The results of technological counteraction to organized crime in Ukraine have been made public. It is about putting an end to the pirated online resources, exposing a fraudulent financial exchange, arresting criminals for the abuse of minors and distribution of relevant content on the closed Internet, suppression of the largest platform for the sale of personal data on the darknet, etc. (Blahuta & Movchan, 2020). There are, however, no classifications of crimes in the investigation of which it is appropriate to use Big Data and Data Mining. This entails a lack of general procedures for the use of information and telecommunication technologies and failure to use all opportunities for international law enforcement cooperation. The last aspect is extremely important, because modern organized crime is transnational. This determines the need for comprehensive support of investigations, in particular, joint investigative teams (European Parliament and the Council of the European Union, 2018).

The procedural aspect is the next problematic issue, that is enshrinement of the results of the use of Big Data and Data Mining in criminal proceedings. The matter is primarily about digital evidence, which is obtained on the basis of processing large information volumes. Conventional analytical methods are not appropriate for managing such data effectively (Usha et al., 2020). Such evidence includes electronic documents (text documents, graphic images, plans, photographs, video and sound recordings, etc.), websites, text, multimedia and voice messages, metadata, databases and other digital information (Blahuta & Movchan, 2020, p. 112).

Crime forecasting is closely related to the problems of criminal investigation. It is extremely difficult to detect crimes and investigate large-scale criminal activities of organized groups without proper organization of analytical work in this area. Forecasting crime is one of the most difficult tasks in law enforcement. In particular, the Big Data method has shown the potential of generalizing such indicators as geography, education, housing availability, urbanization, and population structure to predict the risk of crime in large cities (Wang et al., 2020). Trying to estimate hidden (latent) crime indicators is a separate problem (Jha et al., 2021). This is the reason for the

experts to emphasize the relevance of an intellectual expert system that involves methods of intellectual data analysis to predict the criminogenic situation (Norouzi & Ataei, 2021). In this regard, Data Mining enables combining formalized approach and informal analysis, as well as quantitative and qualitative data analysis (Dehtiarovai & Yevdokimov, 2018).

So, innovations in the investigation of criminal offences, including modern crime forecasting capabilities, allow for a better allocation of law enforcement resources (Hou et al., 2022). Therefore, it is emphasized that the correct use of Big Data and Data Mining can provide significant savings of public funds that are allocated to the field of security (Hassani et al., 2016).

The prospects for the widespread use of Big Data and Data Mining in the investigation of criminal offenses encounter difficulties that can be divided into several groups:

- lack of qualified personnel. The use of Data Mining is affected by the growth of Big Data volumes, but for people who do not have data analysis skills and do not have special knowledge (Hassani et al., 2016) the admissibility of such work for the investigation of criminal offenses is doubtful;
- a certain subjectivity in the selection and assessment of primary data. Detectives and experts still have certain prejudices about the collection and analysis of DNA, fingerprints, electronic messages, etc. (Oatley et al., 2020);
- the impact of the latency of crimes on the formation of databases, which leads to an inadequate analysis of the criminal situation (Guariglia, 2020);
- a time factor affecting the reliability of the results of using the latest methods. For example, the use of Data Mining is effective, but mostly in small time intervals (Dehtiarovai & Yevdokimov, 2018).

Along with this, there is a danger of violation of human rights and freedoms during the investigation of crimes using Big Data and Data Mining. For example, facial recognition systems can be used to covertly collect data not only on criminals, but also on citizens who have never been in trouble with the law. Besides, scanning the profiles of social media users gives law enforcement officers access to the private lives of millions of people (Blahuta & Movchan, 2020). This is why some states impose restrictions on

the use of the latest technologies in law enforcement activities. For example, starting in 2020, some cities in the US significantly limited the allocation of resources for policing in accordance with analytics that can predict future crime locations, potential victims, and criminals (Guariglia, 2020).

So, the use of Big Data and Data Mining in the investigation of criminal offences is an urgent problem that has both huge positive prospects and objective difficulties. The legal dimension of this problem draws attention to criminal law, procedural and human rights aspects. It is necessary to settle them for the widespread use of Big Data and Data Mining in the field of criminal justice.

Methodology and methods

The literature that covers the legal, procedural, and technological aspects of using Big Data and Data Mining in the investigation of criminal offences, as well as forecasting criminal activity was selected to achieve the aim set in the article and fulfil its objectives. Their analysis made it possible to identify the main components of the subject under research, which reflect the legal dimension of the problem.

The article also involved a generalization of the practice of international law enforcement organizations regarding the results of the use of Big Data and Data Mining in the field of criminal justice in terms of the requirements for building an evidence base in criminal proceedings. This gave grounds to determine the main prospects for making the application of these methods for the investigation of criminal offences and forecasting of criminal activity more effective.

The aim of the research was achieved through the following methods:

- systemic approach was used to study the tasks and technologies of Big Data and Data Mining in the field of criminal justice in terms of human rights protection;
- descriptive analysis was used to identify the specifics of Big Data and Data Mining as innovative forensic methods;
- systematic sampling and doctrinal approach enabled identifying and describing the features of criminal offences which can be investigated with the use of Big Data and Data Mining;
- forecasting was used to determine the prospects for making the use of Big Data and Data Mining as methods of investigating

crimes and predicting criminal activity more effective.

Results

Innovations in the methods of detection and investigation of criminal offences reflect the intensive use of technologies in criminal activities and the demand for digital evidence in criminal proceedings. The application of Big Data and Data Mining, which enable organizing and using significant arrays of structured and unstructured information, in combating crime is a many-sided problem. It includes:

- a) features of crimes that can be investigated using Big Data and Data Mining;
- b) crime combating objectives that can be fulfilled with the help of these methods;
- c) the specifics of using Big Data and Data Mining methods and technologies in criminal proceedings;
- d) requirements for the application results;
- e) compliance with basic human rights and freedoms.

Defining the range of crimes is complicated by the heterogeneity and number of their types that can be considered in this context. Big Data and Data Mining cover a wide range — from simple theft to international criminal activity. At the same time, information about suspects can be obtained and stored in different countries and cover significant periods of time (Hassani et al., 2016). It follows that the detection of such crimes usually requires cooperation with foreign states and coordination of international organizations, for example, Europol or Interpol. The conceptual documents contain only an approximate list of such crimes, for example, in Annex 1 to the Regulation (EU) 2018/1727 of The European Parliament and of The Council (2018). It is considered that the relevant criminal offences can be classified according to the following criteria: a) territorial affiliation; b) the nature of the act; c) subject composition. At the same time, classification groups do not exclude each other, but describe actions in different aspects. The following can be considered as the main common feature of all acts: a) the complexity of their investigation, which necessitates the use of the latest technologies; b) dangerousness, as a result of which they are classified as serious (serious) crimes punishable by imprisonment (see Figure 1).

As already mentioned, Big Data and Data Mining are methods that are used not only in the investigation of criminal offenses, but also in

predicting crime. These two tasks are closely related, because the element of delinquency is crime. In this context, crime forecasting should be considered as a logical operation with a purpose of identifying and, as a result, investigating particular criminal offences. The regularities that are revealed through the analysis of crime are also of significant importance at the level of a separate crime. The prediction of connections is a key area of research in complex social systems, which can be implemented by assessing the possibility of the non-obvious connections between pairs of objects. This can provide an effective means of detecting hidden connections in criminal networks and conspiratorial criminal groups (Assouli et al., 2021).

Figure 2 shows the relationship between crime investigation and crime forecasting in the context of Big.

Data and Data Mining

A large number of data sources with the entire set of structured and unstructured data contained cause the urgent need to use Big Data and Data Mining for crime forecasting and crime investigation. These can be open sources and those that require permission to work with them. They can belong to the state, law enforcement agencies, commercial entities, public organizations, individuals (Blahuta & Movchan, 2020). They are the source material to be processed by law enforcement agencies through Data Mining to create various data bases (banks) that are actively used in the crime investigation. This is especially important for international investigations. For example, the General Secretariat of Interpol has created and operates data banks that contain information:

- a) about persons wanted for crimes, missing persons, persons subject to identification, in particular, unidentified corpses, etc.;
- b) about vehicles stolen on the territory of Interpol member states;
- c) about stolen/lost identification documents, as well as stolen/lost forms of administrative documents;
- d) about works of art, antiques, other cultural values stolen on the territory of Interpol member states;
- e) about DNA recovered from crime scenes on the territory of Interpol member states and from criminals;
- f) fingerprints recovered from crime scenes on the territory of Interpol member states and from criminals;

- g) that enables identification of pornographic images;
- h) a bank of pornographic images created with the involvement of minors;
- i) a bank of images of counterfeit payment cards and their elements, as well as other relevant information regarding forgery of payment cards, etc. (Interpol, n.d.).

Figure 3 presents the general classification of the most popular data sources and their relationship with databases. At the same time, it does not contradict the appropriateness of analysing data from many sources (Multi-Source Analysis) (Blahuta & Movchan, 2020).

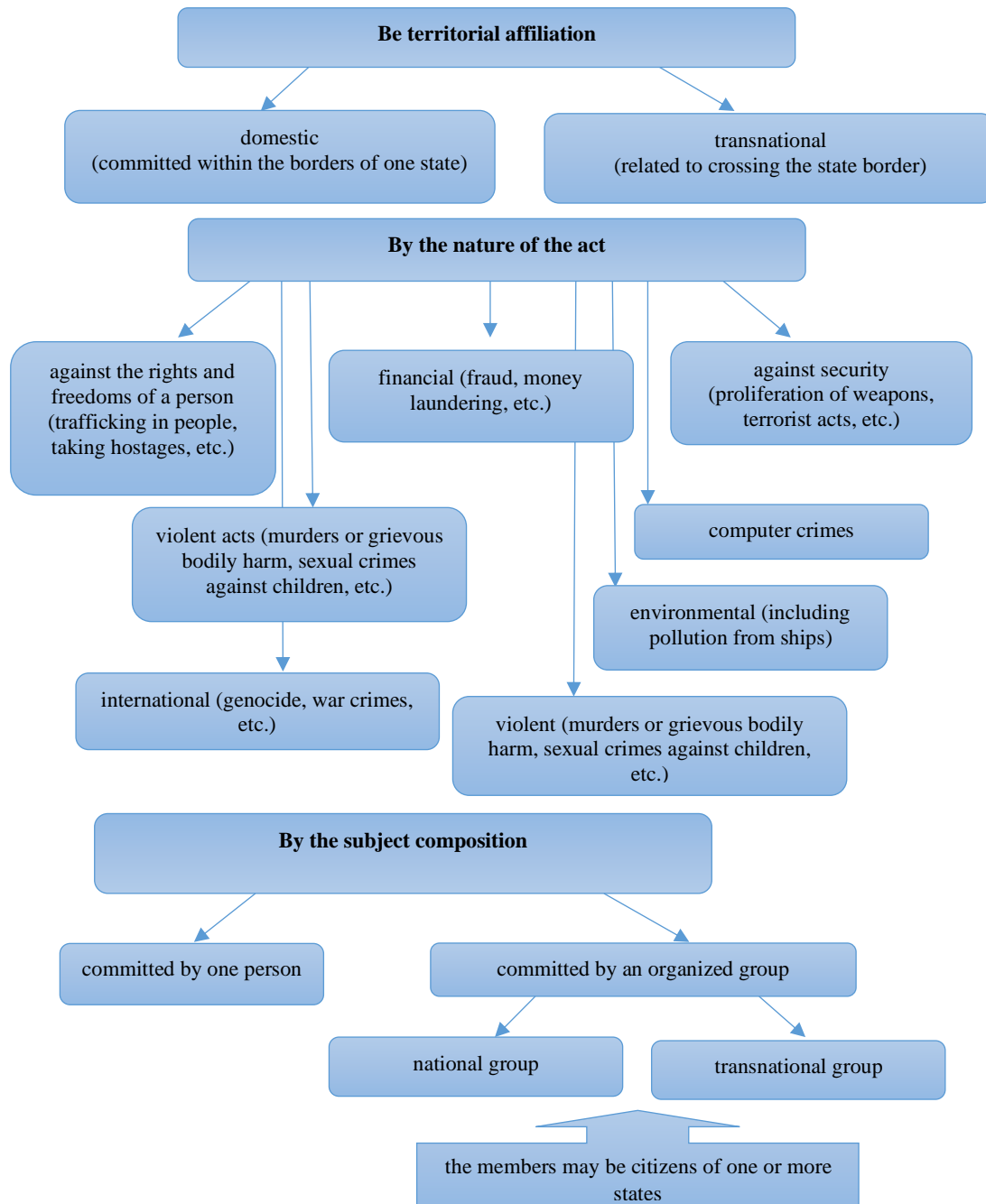


Figure 1. Classification of crimes which can be investigated using Big Data and Data Mining (built based on European Parliament and of The Council (2018))

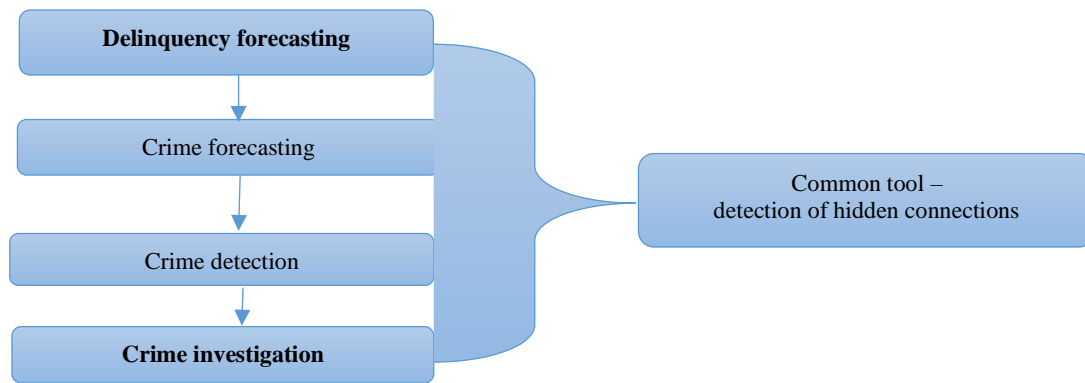


Figure 2. Crime forecasting and investigation as a Big Data and Data Mining task (developed by authors)

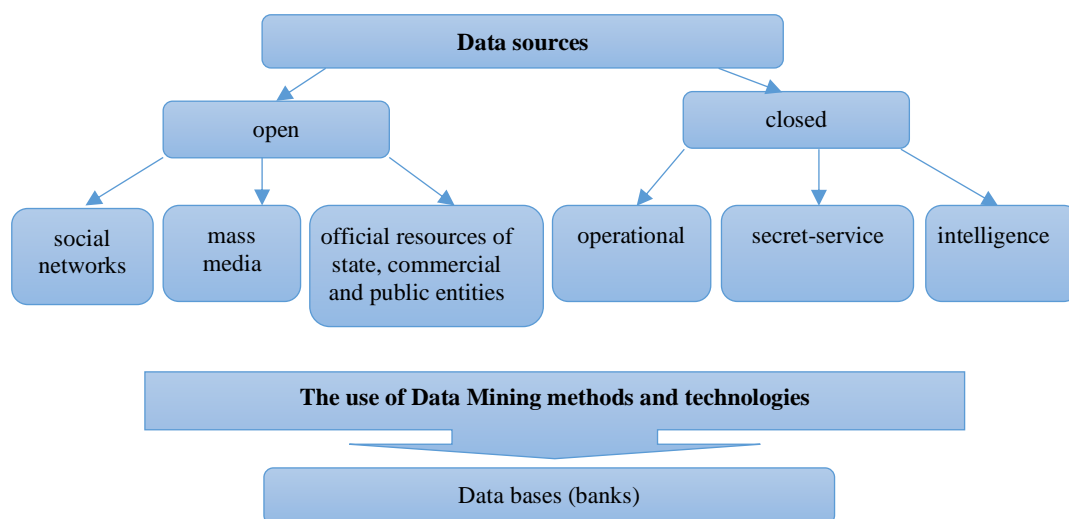


Figure 3. Classification of data sources and their relationship with data bases (banks) (based on Blahuta and Movchan (2020))

Technologically, Big Data are processed using Data Mining methods and technologies implemented through computer tools. A certain combination of methods is determined by the analyst taking into account the task and specifics of the criminal offense under investigation. While classification is the most popular method of Data Mining in analysing delinquency (Hassani et al., 2016), the most popular methods in criminal proceedings are: a) pattern identification; b) cluster analysis (clustering); c) association analysis; d) classification; e) social network analysis. In turn, visualization and machine learning are the main technologies with which Data Mining methods are implemented. Visualization is used to find exceptions, general trends and dependencies, helps in obtaining data at the initial stage of a particular project. Machine learning is further used to find dependencies in the project that has already been launched (Dehtiarovai & Yevdokimov, 2018). The specifics of the main Data Mining methods in

relation to criminal proceedings are presented in Figure 4.

The results of the application of Data Mining methods and technologies in criminal proceedings must be subject to certain requirements. They are determined by the tasks of certain components of the work of investigators and experts. Data Mining creates conditions for minimal user intervention in obtaining results. This is useful for analysts and practitioners to make important decisions (Norouzi & Ataei, 2021) regarding crime investigations. In particular, it is about the objectification of the process of advancing versions, the choice of tactical methods of investigative actions, etc. The exchange of information between law enforcement agencies is an important component of work with Big Data and Data Mining. The specifics of crimes for which Big Data and Data Mining are used for investigation sets a requirement that the results of

the application of these methods facilitate communication between law enforcement officers of different states (Europol, 2022). In general, evidence in criminal proceedings, which

will be recognized in court as admissible, reliable and sufficient for deciding a case on its merits should be the main result.

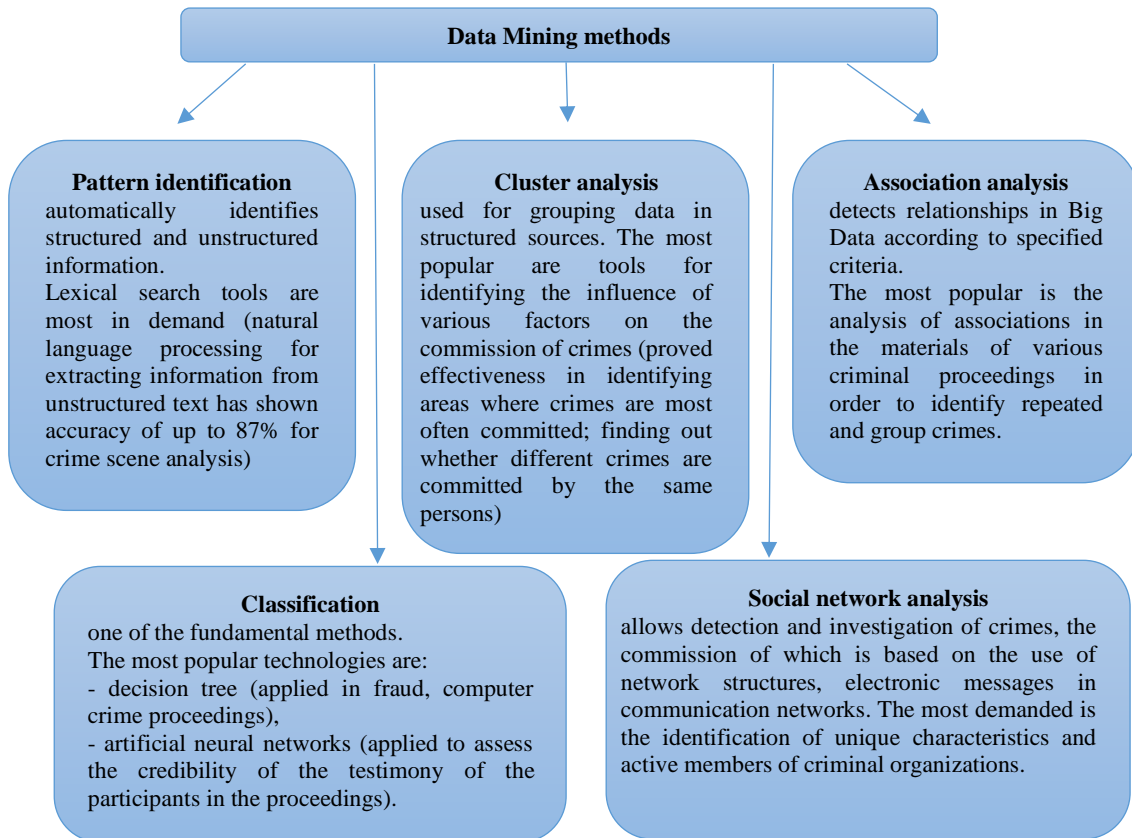


Figure 4. The main Data Mining methods in the context of the investigation of criminal offences (Hassani et al., 2016)

In this context, human rights issues are of particular importance. The use of forensic innovations in criminal proceedings determines the discourse on the provision of human rights and freedoms both in relation to participants in criminal proceedings and in relation to persons whose interests may be affected by the investigation.

The privacy and personal data protection are among the most vulnerable areas. Access to data plays an important role in the effectiveness of Data Mining as a forensic method. However, the need to keep the information confidential causes problems (Hassani et al., 2016). In particular, regulatory acts regulate this issue at the EU level. It is noted that personal data must be processed in a legal, fair and transparent manner in relation to the data subject; such data must be relevant and limited to the purposes for which they are

collected; they must be stored in such a way as to ensure the security of personal data, including their protection from unauthorized or illegal processing (European Union, 2018). Controversial issues of excessive interference in the private life of vulnerable categories of persons (children, the elderly, persons in need of international protection, etc.) may arise during the processing of personal data in criminal proceedings. Such issues must be resolved with full respect for human dignity and integrity (European Parliament and the Council of the European Union, 2019). The use of this method of Data Mining as analysis of social networks urges the issue of freedom of speech and expression on the Internet (Guariglia, 2020), etc. Figure 5 summarizes risks of violation of human rights and freedoms caused by the use of Big Data and Data Mining in the investigation of crimes.

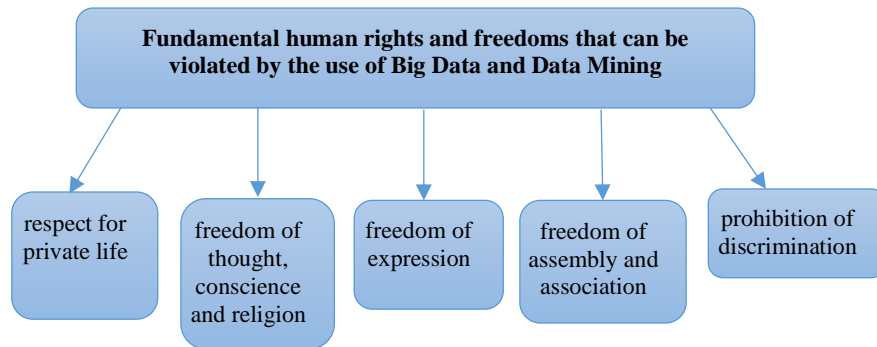


Figure 5. The main risks of using Big Data and Data Mining in criminal proceedings in the context of human rights protection (based on European Union (2018), European Parliament and the Council of the European Union (2019))

The mentioned aspects enabled determining the main prospects for improving the use of Big Data and Data Mining in the investigation of criminal offences, including in crime forecasting. They

are seen to be related to the standardization of procedures for their use. The following principles should be the basis of these standards: a) ethical; b) organizational; c) procedural (see Figure 6).

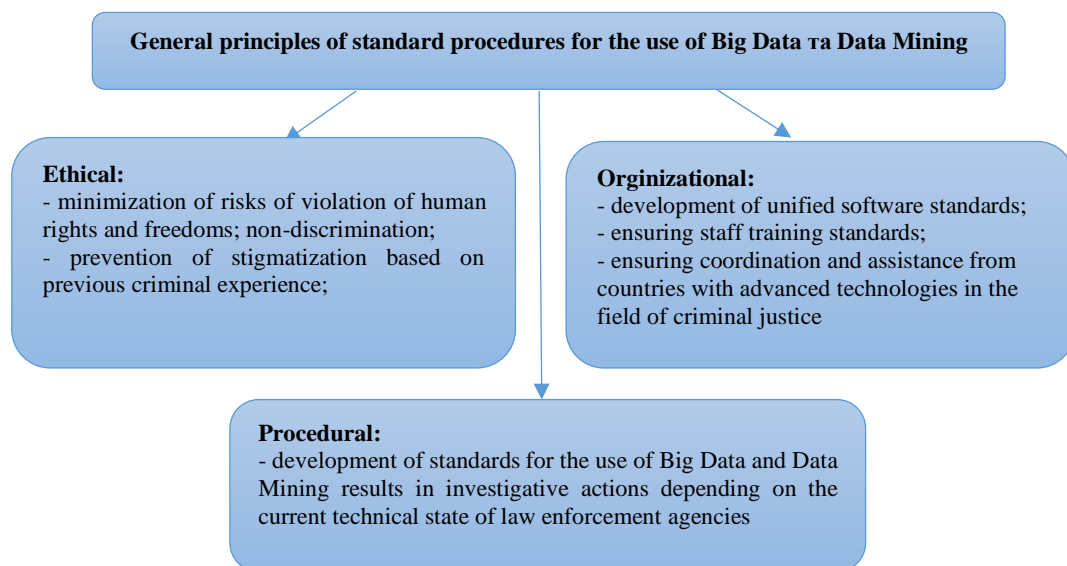


Figure 6. Prospects for improving the use of Big Data and Data Mining in the investigation of criminal offences (developed by authors)

In view of the above, it is considered appropriate to talk about the development of standard procedures for the use of Big Data and Data Mining in the investigation of criminal offences. These procedures should be based on ethical, organizational and procedural principles. It is appropriate to set out the relevant framework procedures in practical recommendations for authorized persons of law enforcement agencies, noting that violation of their principles will entail responsibility. This will enable to actively apply Big Data and Data Mining in criminal proceedings and use the results for the needs of national and international justice.

Discussion

Studies on the use of Big Data and Data Mining in the investigation of criminal offences is mainly focused on the software features, algorithms and their improvement. The legal aspects of the problem are much less studied. However, in general, the professional discussion is conducted in the context of the appropriateness and effectiveness of using the latest technologies in combating criminal activity. The focus is if the issue of crime forecasting, much less often — forecasting the commission of individual crimes. In general, the interrelationship of these aspects

for the fulfilment law enforcement objectives is not studied.

One should agree with the general initial thesis that the evolution of criminal behaviour led to the use of the latest technologies not only to commit crimes, but also to avoid punishment (Hassani et al., 2016). In this regard, Big Data and Data Mining, which are aimed at identifying accurate, simple, expedient and understandable patterns and models (Belesiotis et al., 2018), enable automating the detection of patterns and relationships in large data sets regarding crime. It is rightly noted that these methods play a significant role in informational support for the decision-making regarding crime control and crime prevention (Norouzi & Ataei, 2021).

In this study, the position regarding the prevalence of open data sources over closed ones was confirmed, because law enforcement agencies receive from 35% to 95% of data from open sources (Blahuta & Movchan, 2020). At the same time, it is debatable that the authors mainly focus on certain types of data. In particular, attention is paid to the prospect of reconnaissance into large groups of people using sensor devices, which provides information on the dynamics of social processes (Zhou et al., 2021); much attention is paid to geolocation in the context of proactive decision-making and prevention (Butt et al., 2020; Hussain & Aljuboori, 2022), etc. In this regard, one should agree with the opinion that forecasting crime requires significant improvement in the quality of Big Data, including the analysis of housing prices, population density, traffic conditions, and the unemployment rate (Hajela et al., 2021; Hou et al., 2022) etc. Crime-related events reveal spatio-temporal patterns that can also be used for prediction and subsequent decision-making (Kadar et al., 2019). However, other authors are sceptical about the accuracy of the prediction based on the inadequacy of the processed data (Wang et al., 2020).

But in relation to the investigation of criminal offences, the data obtained in the course of different criminal proceedings require a comprehensive analysis. First of all, the mutual connections of the participants in several investigations are revealed (Blahuta & Movchan, 2020). The analysis of such data as the time of day, season of the year, weather data, types of victims and features of places where crimes are committed is also promising. This helps to conclude when and where crimes are most likely to be committed (Guariglia, 2020). In view of the foregoing, the position should be shared that the

efficiency of Data Mining can be significantly increased only through the combination of data from several sources of information (Belesiotis et al., 2018). Accordingly, an approach that offers a combination of Data Mining methods (Dehtiarovai & Yevdokimov, 2018) focused on structured and unstructured data (Hassani et al., 2016) is promising.

This research confirmed that the use of Big Data and Data Mining encounters objective difficulties. However, one cannot agree on the predominant attention to technological aspects: the formation of clusters (regions) according to the criterion of the average score of the risk of becoming a victim of certain illegal actions (Soni et al., 2019); analysis of data sources on some types of criminal activity with unsatisfactory content quality (Hassani et al., 2016). For some types of crimes, the appropriateness of using the latest technologies is generally doubtful because of the uniqueness of criminal activity, as there is simply not enough data for analysis (Dupont et al., 2018). In such cases, the legal dimension of the problem cannot be covered.

Another problematic aspect is that insufficient attention is paid in the literature to the micro level — a specific criminal offense. In particular, the authors emphasize that although criminals constantly improve their criminal activities, they find new ways of committing crimes and evading social control (Dupont et al., 2018). Therefore, technology cannot predict crime. This only undermines trust between the police and society, discriminates against vulnerable populations and creates a greater risk of crime (Guariglia, 2020). Such views are a certain exaggeration and generally call into question the intellectual methods of investigating criminal offenses.

In contrast, the discourse on the human rights component of the use of Big Data and Data Mining in the investigation of criminal offenses is more realistic. Today, huge data sets can be collected and analysed secretly (Blahuta & Movchan, 2020). The non-transparent data processing tools, which lead to the creation of certain ratings regarding the risk of committing crimes by persons with criminal experience, are subject to sound criticism (Guariglia, 2020). One must, however, agree that these ethical and organizational issues must be addressed before new technologies become widely used and implemented in all criminal justice procedures (Dupont et al., 2018). This thesis was confirmed and developed in the results of our research.

In general, these considerations can be the basis of the legal, organizational and procedural aspects of the application of Big Data and Data Mining in the investigation of criminal offences.

Conclusions

The conducted research gave grounds for drawing a number of conclusions regarding the use of Big Data and Data Mining for the detection and investigation of criminal offences.

It was established that the mentioned criminal offences have different criminal law characteristics. They can be combined into a single group based on two factors — severity (their commission is punishable by imprisonment) and the complexity of the investigation. The features of Big Data and Data Mining make it possible to use them both for crime investigation and crime forecasting, which are interrelated tasks in the field of law enforcement. It is shown that the processing of heterogeneous open and closed data sources through Data Mining enables creating data bases (banks) used in law enforcement activities. The procedure for the use of Data Mining represents the use of methods and relevant procedures. The main Data Mining methods used in investigation of criminal offences, as well as aspects of their most frequent application are shown. It was established that the use of Big Data and Data Mining is associated with the risks of violation of basic human rights and freedoms. The most vulnerable objects of violations in this area were identified. The further implementation of Big Data and Data Mining in criminal proceedings is connected with the standardization of procedures for the use of particular methods or their set.

The standardization of procedures is proposed in order to unify the results of using Big Data and Data Mining in the context of preparing evidence for national and international courts. It is proposed to develop the ethical, organizational and procedural principles of each procedure and present them in practical recommendations for authorized persons of law enforcement agencies. Responsibility for violation of the specified principles shall be a separate aspect of those procedures.

Prospects for further research of forensic innovations in the investigation of criminal offences include standardization of their use to prepare an evidence base in the interests of criminal justice. A separate promising direction is the specialized training of specialists for the development, implementation and use of the latest technologies in criminal proceedings.

Bibliographic References

- Assouli, N., Benahmed, Kh., & Gasbaoui, B. (2021). How to predict crime - informatics-inspired approach from link prediction. *Physica A: Statistical Mechanics and its Applications*, 570, 125795. <https://doi.org/10.1016/j.physa.2021.125795>
- Belesiotis, A., Papadakis, G., & Skoutas, D. (2018). Analyzing and predicting spatial crime distribution using crowdsourced and open data. *ACM Transactions on Spatial Algorithms and Systems*, 3(4), 1-31. <https://doi.org/10.1145/3190345>
- Blahuta, R., & Movchan, A. (2020). The latest technologies in the investigation of crimes: The current state and problems of use. Lviv: Lviv State University of Internal Affairs.
- 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.
- Chaudhary, M., & Bansal, D. (2022). Open source intelligence extraction for terrorism-related information: A review. *WIRES. Data Mining and Knowledge Discovery*, Online version, e1473. <https://doi.org/10.1002/widm.1473>
- Das, P., Das, A. K., Nayak, J., Pelusi, D., & Ding, W. (2021). Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. *Information Sciences*, 562, 279-303. <https://doi.org/10.1016/j.ins.2021.02.002>
- Dehtiarovai, Y. V., & Yevdokimov, Y. (2018). Data mining methods and models for social and economic processes forecasting. *Mechanism of Economic Regulation*, 2, 34-44. <https://doi.org/10.21272/mer.2018.80.03>
- Dupont, B., Stevens, Y., Westermann, H., & Joyce, M. (2018). Artificial intelligence in the context of crime and criminal justice. Montreal university [Université de Montréal]. <http://dx.doi.org/10.2139/ssrn.3857367>
- European Parliament and the Council of the European Union. (2018). Regulation (EU) 2018/1727 of the European parliament and of the council of 14 November 2018 on the European Union Agency for Criminal Justice Cooperation (Eurojust), and replacing and repealing Council Decision 2002/187/JHA. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1727&from=IT>



- European Parliament and the Council of the European Union. (2019). Regulation (EU) 2019/817 of the European parliament and of the council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of borders and visa and amending Regulations (EC) No 767/2008, (EU) 2016/399, (EU) 2017/2226, (EU) 2018/1240, (EU) 2018/1726 and (EU) 2018/1861 of the European Parliament and of the Council and Council Decisions 2004/512/EC and 2008/633/JHA. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019R0817>
- European Union. (2018). The General Data Protection Regulation: Regulation (EU) 2016/679. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>
- Europol. (2022). Secure Information Exchange Network Application: Ensuring the secure exchange of sensitive and restricted information. Retrieved from <https://www.europol.europa.eu/operations-services-and-innovation/services-support/information-exchange/secure-information-exchange-network-application-siena>
- Grechkina, O., Korniyushkina, A., Naruzhnaya, E., Tonkov, E., & Turanin, V. (2019). El lenguaje jurídico como medio de comunicación intelectual y jurídico. *Revista Científica Del Amazonas*, 2(3), 32-38. Recuperado a partir de <https://revistadelamazonas.info/index.php/amazonas/article/view/15>
- Guariglia, M. (2020). Technology can't predict crime, it can only weaponized proximity to policing. Electronic Frontier Foundation. Retrieved from <https://www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing>
- Hajela, G., Chawla, M., & Rasool, A. (2021). A multi-dimensional crime spatial pattern analysis and prediction model based on classification. *ETRI Journal*, 43(2), 272-287. <https://doi.org/10.4218/etrij.2019-0306>
- Hassani, H., Huang, X., Silva, E. S., & Ghods, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining*, 9(3), 139-154. <https://doi.org/10.1002/sam.11312>
- Hou, M., Hu, X., Cai, J., Han, X., & Yuan, S. (2022). An integrated graph model for spatial-temporal urban crime prediction based on attention mechanism. *ISPRS International Journal of Geo-Information*, 11(5), 294. <https://doi.org/10.3390/ijgi11050294>
- Hussain, F. S., & Aljuboori, A. F. (2022). A crime data analysis of prediction based on classification approaches. *Baghdad Science Journal*, 5, 1073-1077. <http://dx.doi.org/10.21123/bsj.2022.6310>
- Interpol. (n.d.) Our 19 databases. Recovered from <https://www.interpol.int/How-we-work/Databases>
- Jha, S., Yang, E., Almagrabi, A. O., Bashir, A. K., & Joshi, G. P. (2021). Comparative analysis of time series model and machine testing systems for crime forecasting. *Neural Computing and Applications*, 33, 10621-10636. <https://doi.org/10.1007/s00521-020-04998-1>
- Kadar, C., Maculan, R., & Feuerriegel, S. (2019). Public decision support for low population density areas: An imbalance-aware hyperensemble for spatio-temporal crime prediction. *Decision Support Systems*, 119, 107-117. <https://doi.org/10.1016/j.dss.2019.03.001>
- Norouzi, N., & Ataei, E. (2021). Application of data mining in identifying and discovering hidden patterns of theft. *International Journal of Innovative Research in the Humanities*, 1(1), 29-42.
- Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. *WIREs Data Mining and Knowledge Discovery*, 12(2), e1432. <https://doi.org/10.1002/widm.1432>
- Oatley, G., Chapman, B., & Speers, J. (2020). Forensic intelligence and the analytical process. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1354. <https://doi.org/10.1002/widm.1354>
- Pokhriyal, N., Kumar, N., Verma, R., & Semwal, A. (2020). Survey on crime data analysis using a different approach of K-Means clustering. *International Journal of Advanced Science and Technology*, 29(5), 13839-13854.
- Pramanik, M. I., Lau, R. Y. K., Yue, Wei T., Ye, Y., & Li, C. (2017). Big data analytics for security and criminal investigations. *WIREs Data Mining and Knowledge Discovery*, 7(4), e1208. <https://doi.org/10.1002/widm.1208>
- Soni, S., Shankar, V. G., Chaurasia, C. (2019). Route-the safe: A robust model for safest route prediction using crime and accidental data. *International Journal of Advanced Science and Technology*, 28(16), 1415-1428.
- Usha, D., Niveditha, V. R., Kirubadevi, T., & Thamizhikkavi, P. (2020). Use of predictive analytical algorithm by crime investigation

- team – An Analysis. *International Journal of Advanced Science and Technology*, 29(9s), 2986-2992.
- Wang, J., Hu, J., Shen, S., Zhuang, J., & Ni, S. (2020). Crime risk analysis through big data algorithm with urban metrics. *Physica A: Statistical Mechanics and its Applications*, 545, 123627. <https://doi.org/10.1016/j.physa.2019.123627>
- Zhao, X., & Tang, J. (2017). Modeling temporal-spatial correlations for crime prediction. In: E.-P. Lim, & M. Winslett (Eds.), *Proceeding of the 2017 ACM on Conference on Information and Knowledge Management*, (pp. 497-506). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3132847.3133024>
- Zhou, B., Chen, L., Zhao, S., Zhou, F., Li, S., & Pan, G. (2021). Spatio-temporal analysis of urban crime leveraging multisource crowd sensed data. *Personal and Ubiquitous Computing*. Retrieved from <https://doi.org/10.1007/s00779-020-01456-6>

