Ceza Hukuku ve Kriminoloji Dergisi-Journal of Penal Law and Criminology 2020; 8(2):293-319
ISSN: 2148-6646 / E-ISSN: 2602-3911

http://jplc.istanbul.edu.tr

Research Article / Araştırma Makalesi

DOI: 10.26650/JPLC2020-813328

Big Data, Data Mining, Machine Learning, and Deep Learning Concepts in Crime Data

Emre Cihan ATEŞ[1], Erkan BOSTANCI[2], Mehmet Serdar GÜZEL[2]

[1]Gendarmerie Captain, Lecturer, Department of Security Science, Gendarmerie and Coast Guard Academy, Ankara, Turkey
[2]Ph.D., Assistant Professor, Department of Computer Science, Ankara University, Ankara, Turkey

ORCID: E.C.A. 0000-0001-9550-4532; E.B. 0000-0001-8547-7569; M.S.G. 0000-0002-3408-0083

ABSTRACT

Along with the rapid change of information technologies and the widespread use of the internet over time, data stacks with ample diversity and quite large volumes has emerged. The use of data mining is increasing day by day due to the huge part it plays in the acquisition of information by making necessary analyses especially within a large amount of data. Obtaining accurate information is a key factor affecting decision-making processes. Crime data is included among the application areas of data mining, being one of the data stacks which is rapidly growing with each passing day. Crime events constitute unwanted behaviour in every society. For this reason, it is important to extract meaningful information from crime data. This article aims to provide an overview of the use of data mining and machine learning in crime data and to give a new perspective on the decision-making processes by presenting examples of the use of data mining for a crime. For this purpose, some examples of data mining and machine learning in crime and security areas are presented by giving a conceptual framework in the subject of big data, data mining, machine learning, and deep learning along with task types, processes, and methods.

Keywords: Crime, big data, data mining, security, policing

Submitted: 20.10.2020 • **Revision Requested:** 15.12.2020 • **Last Revision Received:** 19.12.2020 • **Accepted:** 22.12.2020 • **Published Online:** 19.01.2021

Corresponding author: Emre Cihan Ateş, E-mail: emre_cihan_ates@hotmail.com

Citation: Ates EC, Bostanci E, & Guzel MS, 'Big Data, Data Mining, Machine Learning, and Deep Learning Concepts in Crime Data' (2020) 8(2) Ceza Hukuku ve Kriminoloji Dergisi-Journal of Penal Law and Criminology, 293. https://doi.org/10.26650/JPLC2020-813328

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

1. Introduction

In order to examine the concept of data mining, first the concepts of data, knowledge, information, and wisdom should carefully be examined. In this context, the term "data" is expressed in the form of definitions such as knowledge that is hidden but not yet interpreted or analysed (Zins, 2007), simple observations and symbols, raw/unorganized facts (Ahmed, 2020), and a text that does not answer certain questions (Shao et al., 2017).

The term "information" is defined using expressions such as the valuable knowledge in the human mind, contextual information including experience, values, predictions, a collection of data arranged in a consistent way (Ahmed, 2020; Cooper, 2017), and a text that answers why-how questions and is expressed as an output derived from data and strained through the mind. The word "knowledge" refers to the message that changes the knowledge level and purpose of a person, the perception of its recipient, a text that answers questions such as who, when, what, or where, and data that makes sense. Wisdom, on the other hand, involves a foresight for the future taking advantage of existing knowledge. Figure 1 gives an illustration of these concepts in the form of a pyramid.

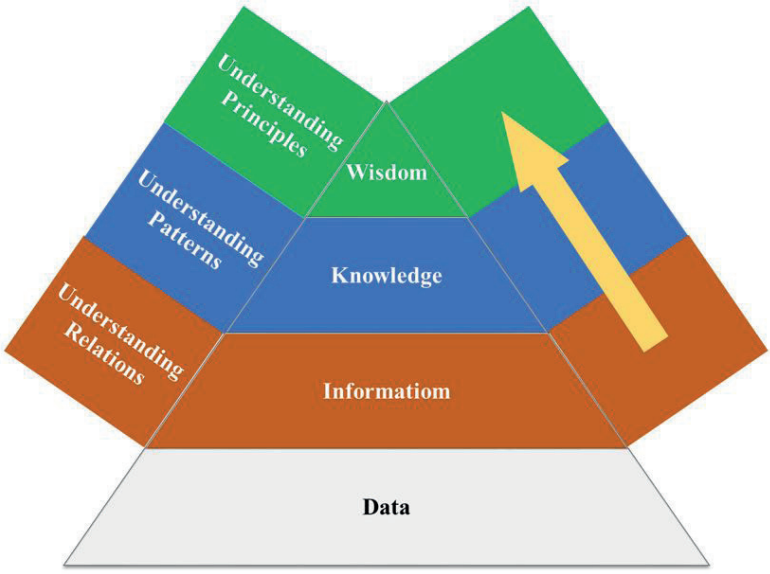


Figure 1. Data, information, knowledge, and wisdom

In the pyramid shown in Figure 1 (Hey, 2004; Cooper, 2017),

- The concept of data expresses a rather general concept that is not yet analysed,
- The term information refers to the straining of data through the mind,
- The concept of knowledge expresses a form of data that has become meaningful.

The concept of wisdom comes from a combination of the steps of data, information and knowledge. Together with the concept of wisdom, that of big data comes out raw at first, but becomes meaningful by being processed and thus helps our decision making activities (Pauleen et al., 2017).

The term 'big data' is used to describe big data at every level, whether it is raw or processed. Data is like precious metals waiting to be removed from piles of stone. We live in a world where a lot of data is being collected every day. Parallel to the development of technology, it is also a known fact that data sizes and types have increased as computer and internet concepts have entered our lives. A particular problem that this increase in data volume brings is that of complexity, so meaningful analysed data correctly is becoming more important with each passing day.

As a result of the fact that the process of obtaining meaningful data by extracting unqualified ones from the raw data restricted statistical researchers and the current statistical methods were insufficient for processing big data, the concept of data mining has emerged. In other words, data mining is a systematic approach for analysing and identifying different patterns, and relations within big data (Blei & Smyth, 2017). One of the important issues in the method is the analysis of existing data rather than the collection of a lot of data. Data mining aims to develop prediction models about events that are predicted to appear in the future and to support decision-making processes by taking advantage of the past sample in general (Kelleher & Tierney, 2018). The concept of data mining is also referred to in the literature as "knowledge extraction", "knowledge discovery", "data archaeology", "data/pattern analysis", "knowledge mining", and "information harvesting" (Khare & Shrivasta, 2018; Koyuncugil & Özgülbaş, 2009).

The information technologies and the Internet have significantly altered the field of crime and security (Akdemir & Lawless). Current statistics show that crime is on the rise all over the world. For this reason, the concepts of crime and criminality have steadily increased in parallel with an increase in the amount of big data. Considering that crime is an unwanted behaviour in societies, data and knowledge obtained from

data have become important in determining crime prevention strategies. Determining future goals and strategies only depends on up-to-date, accurate and reliable data.

It is the aim of this review article to give a perspective on decision making processes with the use of data mining in crime data by providing an overview of data mining and analysing data mining methods. After the introduction, this study defines big data, data mining, machine learning, and deep learning concepts respectively. In the "Data mining methods and models" section the concepts are defined, the data mining process is explained, and data analysis models are analysed. The section "Use of big data, data mining, machine learning and deep learning in criminal and security areas" summarizes several cases which are given from the field of security dealing with the struggle against crime, machine learning, deep learning and the use of data mining. In the discussion section, the use of data mining, machine learning and deep learning theories in combating crime is examined. In the "Conclusion" section the concepts of big data and data mining are stated to be actively used in the struggle against crime, and their future significance is emphasized.

2. Background

2.1. Big Data

Given that the concept of "big data" is relatively new, it should first be examined in terms of definitions.

· The concept of big data is a fast-growing and abstract interdisciplinary concept offering potential for growth in every social area, which is being stored at an unprecedented scale (Chen et al., 2014),

· Large volume data mostly comes in very rapidly from various sources such as internet and computer networks,

· It is defined as a set of data that is difficult to access through standard information technologies, and that involves the use of data capture, processing, collection, display, and analysis methods on large data sets (Vaidhyanathan & Bulock, 2014).

In order to further investigate the concept of big data, it will be helpful at this point to examine the principles in the literature which define big data and which are typically called the 'five V' (5V - Volume, Velocity, Variety, Value, Veracity). (Abdullah et al., 2015; Ateş et al., 2020; Wamba et al., 2015; White, 2012);

· Volume: This term is used to describe a constantly increasing amount of data. Along

with the increase of the data, a mass of knowledge has also started to be formed and it becomes more difficult to reach qualified knowledge with such an increase in volume.

Variety: The majority of data produced is produced in non-structural environments such as computers, social networks, mobile phones, and tablets. Data sources are relatively new and new sources are being added day by day. This will increase the number of sources for the data, the data will be in different formats and include different variables after data collection, and this will make the analyser's job more difficult.

· Velocity: This expresses the growth rate of the data, and when this is high it is a sign of the data being big data.

· Veracity: This is an important variable guaranteeing that the data is secure during its flow. It is important to protect it under high-level security measures and not to change it by unauthorized interventions.

· Value: This refers to the evaluation of big data and to the help that it provides to decision support units to gain meaningful information by transforming raw data into information and knowledge

2.2. Data Mining

The concept of data mining is defined in many sources, some of which are given below:

· Uncovering potentially useful knowledge about previously unknown data in a non-confidential way (Srinivas et al., 2010),

· It enables obtaining knowledge transformable into comprehension, understanding, and action by combining statistics with concepts, tools and algorithms, machine learning and the analysis of very big data sets (Williams, 2009).

· Data mining is a step in the entire knowledge discovery process that can be described as a knowledge retrieval or mining extraction process from a large amount of data and is a form of knowledge discovery that is particularly needed to solve problems in a certain area (Beniwal & Arora, 2012; Olson & Lauhoff, 2019).

From these definitions, it is possible to make the following definition taking into account the relationship: Data mining is a process that aims to retrieve knowledge making use of the big data in the past and using methods such as statistics, machine learning, and artificial intelligence, and to use that knowledge in the decision support activities related to the future (Campbell & Ying, 2011; Hand & Adams, 2014).

2.3. Machine Learning

The introduction into our lives of the concept of machine learning started with the question " Can machines think?" first put forward by Alan Turing (Turing, 1950). The concept of machine learning is the ability of a computer to learn by itself, using existing data and experiences. In machine learning, there are different learning methods according to the labelling of the training data (Aggarwal, 2018) In supervised learning, the available data is known as training data, which is the basic input given to the machine to learn the model. The computer attempts to learn the training data with the help of various machine learning algorithms. Determining how successfully the information is learned is assessed with test data. The most important difference between test data and training data is whether the data is predicted or classified by the machine. Unsupervised learning is machine learning where there is no training data and all learning is done according to the similarities and differences between the data. Semi-supervised learning, on the other hand, is the learning realized through a mixed structure, in which there is both training data and unlabelled data.

The concept of machine learning is used for data analysis as well as for data mining, and additionally, it is the artificial intelligence type of the analysis that defines and recognizes patterns using artificial intelligence (Mcclendon & Meghanathan, 2015).

· Predictions can be made by learning data through machine learning, which is a sub-branch of artificial intelligence.

· In machine learning, it is a basic principle that the created system performs the learning process by itself.

· The most important difference between computers and humans is that humans learn from the events they lived in the past which they call experience, while machine learning aims to reach this point through a decision mechanism.

· Providing learning by means of automation-enhancing training data in knowledge access enables efficient automated techniques to be carried out with little human power (Jackson, 2002).

· In essence, machine learning is a learning method of a computer system via sampling, and there are many different machine learning algorithms in various problems or data types.

When the concepts of data mining and machine learning are examined at a basic level, the use of artificial intelligence leads to the conclusion that data systems obtain

knowledge from raw data without intervention (Kelleher & Tierney, 2018). In addition, data mining primarily provides meaningful data retrieval from data stacks. It also enables a minimization of human labour and an optimization of results through the concept of machine learning. For this reason, it is possible to say that data mining and machine learning can be actively used in the process of retrieving knowledge in all fields of science, especially in the natural sciences.

2.4. Deep Learning

Deep learning is machine learning, one of the most popular topics today, and is a field of study that includes artificial neural networks with one or more hidden layers and similar machine learning algorithms. Deep learning consists of the structures that can learn complex structures with multi-layered neural networks and that, through hidden layers, can more easily and successfully learn models that classical machine learning algorithms cannot easily learn (Dey, 2016).

Hidden layers are part of a structure in which mathematical functions and calculations take place in order to obtain the desired output. The number of its layers may vary. Although deep learning is a new approach, it has been actively used in many areas such as semantics, transfer learning, natural language processing, visualization, and crime investigation.

As the number of layers increases, the hardware required for the learning process (primarily the graphics processing unit (GPU) requirement) and the time (due to the large number of parameters) will also increase. In addition, although it is known that deep learning requires much more data than classical algorithms during training, this is not a problem today due to the data abundance we have. However, if the amount of data to be analysed is small, classical algorithms may produce more successful results in many ways.

While classical machine learning algorithms use methods based on statistical analysis to recognize the pattern, there is a modelling similar to the neurons of the human brain in deep learning. For this reason, hyperparameter (tuning) optimization that minimizes loss function can be done in many different ways. With deep learning, a good classification performance is obtained for text, audio, and visual data without human intervention, especially due to the ability to extract features.

3. Data Mining Methods and Models

Data mining is very popular nowadays and it is a known fact that a new method or algorithm is added to literature every day, considering both the rapid change in the structure of information and the increase in the information needs of people. And this is an indication of that the data mining methods and algorithms are moving dynamically rather than staying static. So, this is an indication of the fact that the data mining methods and algorithms are moving dynamically rather than generally remaining static. This is because the methods used to obtain the right knowledge can also change according to the state of the data given that data source structures change. Data mining focuses on three main sections. The classification made according to the editability of the data is given below (Agrahari & Rao, 2017; Ge et al., 2017).

- Structured
- Unstructured
- Semi-structured

Structured data defining an explicitly specified cluster of data play an important role in data analysis at all times due to the convenience of their classification (Kumar & Nagpal, 2019). Most of the biometric verifications used in definitions in the field of security are based on structured data. Unstructured data do not have a pre-defined data model. They cannot be classified and they show variance. 90% of all data are considered to be unstructured data, which reveals the fact of a bigger cluster of data that is hard to analyse (Tirgari, 2012). This is particularly true of social network sites, which are becoming more frequently used day by day, and which evidently play a significant role in unstructured data. Unstructured data improve the ability to obtain a greater amount of information from clusters of data. However, the accuracy of the data obtained may not be as great as the structured data. Semi-structured data help the partial classification and definition of information. It is particularly e-mails, document forming languages, and Web and NoSQL database utilization which are among the most common samples of semi-structured data.

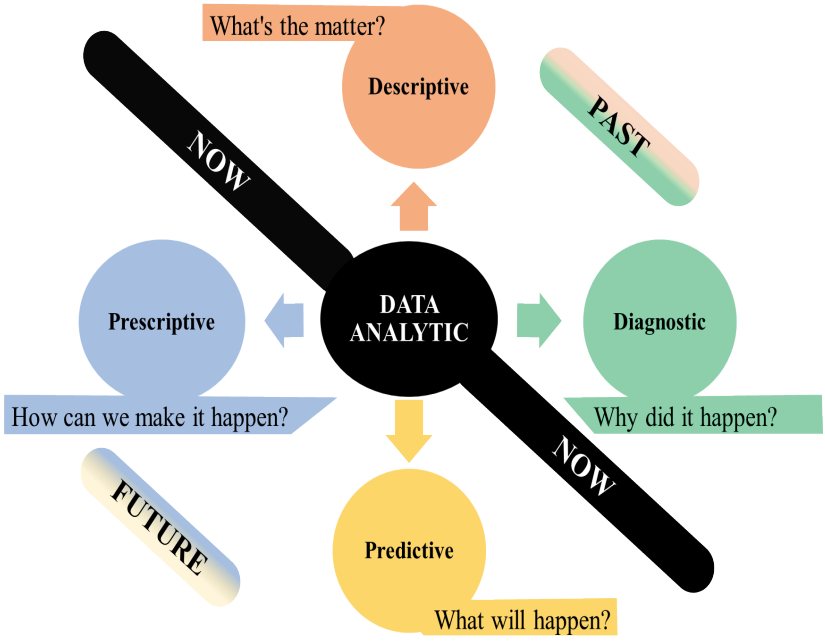


Figure 2. Data Analytic Methods

Data mining, which is examined in three groups as structured, unstructured and semi-structured, is evaluated in four model structures under the general headings of predictive, descriptive, diagnostic and prescriptive (Shown in figure 2) (Mesgarpour & Dickinson, 2014);

· Descriptive Methods: These include methods that determine the links and relationships among the data which will support decision making in the data stack (Bock et al, 2019),

· Diagnostic Methods: These are methods that give reasons and provide knowledge on "why" questions,

· Predictive Methods: These include the methods used to estimate the dependent variable using the independent variables in the database. These are used to make a prediction of future events utilizing known past results,

· Prescriptive Methods: These methods are used to make a prediction of future events utilizing known past results.

If we assert that the majority of the world's data is unstructured, we can easily say that the most commonly used methods are descriptive and diagnostic. Descriptive, diagnostic, predictive and prescriptive methods are included in a sub-hierarchy of structured, unstructured and semi-structured methods. Descriptive, diagnostic, predictive and prescriptive methods are shown in the highest ordering in the hierarchy of methods in many sources since they classify the data mining methods.

The data mining models were examined under the following four main headings according to their functions:

Classification: In classification, categorization of data is usually at the forefront, and this is how data orientations are decided (Mukhopadhyay et al., 2013; Rutkowski et al., 2020). For example, a classification model on crime data can state whether the city is safe or dangerous based on the intensity of crime events in the city. Especially in supervised learning within the scope of classification, it is possible to estimate the classification of recently added issues by making a pattern discovery from the data. Many operations, such as voice recognition, call forwarding, and text classification can be carried out with these algorithms, especially genetic algorithms, the classification of which is rule-defined, that is, supervised. Naive Bayes, logistic regression, genetic algorithms, support vector machines, k-NN (k-nearest neighbours algorithm) and memory-based reasoning are commonly used algorithms, and fuzzy logic is also frequently used.

Regression: This is the method of analysis that can set up a model by predicting data orientations. Unlike classification, it takes place as a prediction that is essential in regression (Mittal et al., 2019). For example, the regression model can be used to predict potential future data for crimes by analysing the current types, times and frequencies of crimes in the city.

Clustering: This is a process of decomposing into groups called "clusters" based on a certain proximity criterion (Berkhin, 2006; Gupta & Chandra, 2019). It is expected that the data in the same cluster will be similar to each other and that the similarity will be much smaller in different clusters in general after the clustering process has been performed (Rutkowski et al., 2020). Clustering, in other words, grouping, has become even more important especially with the concept of machine learning. It is actively used in many areas such as voice and image processing, speech recognition, frequent phone calls, messaging and data usage, customer sorting, and especially for

the purpose of social network analysis. Korhonen Artificial Neural Networks, Canopy, Mean Shift, K-means, Fuzzy C-averages, Latent Dirichlet Allocation, k-medoids and MinHash are examples of clustering algorithms which are in use (Mukhopadhyay et al., 2013; Ghorbani & Ghousi, 2019).

Association rules: These are used to support future studies by determining the associated behaviours and making use of the data obtained from the past (Ngai et al., 2009, Mittal et al, 2019). It is particularly the revealing of attitudes encountered in organized and frequently committed crimes which will enable detection to be done much faster.

4. Use of Big Data, Data Mining, Machine Learning and Deep Learning in Criminal and Security Areas

Data mining and machine learning concepts have recently become more popular as they are easier to use through bundled software. This popularity has also increased the usage areas of data mining. As the stack of data to be examined grows, the importance of concepts such as data mining, artificial intelligence, and machine learning has once again shown itself. As long as the population of humankind grows and rules exist in the world, crime data will continue to increase, just like other data stacks.

The concept of crime refers to the practice of activities which are prohibited in accordance with norms and deviations in societies. Although it varies from society to society, the concept of crime is generally defined as acting contrary to existing laws. It is possible to put types and kinds of crime into categories, from sexual crimes to public order crimes, from human trafficking to drug-based crimes, from juvenile delinquency to war crimes, and these will push the limits of imagination of human beings. In this case, the fight against crime is becoming more important for societies for the continuation of the existing social order.

Law enforcement applications in the fight against crime are generally examined under two headings, namely preventive and reactive (in other words, judicial) (Clarke, 2006). Reactive crime investigations aim to identify elements such as unknown perpetrators, or the victim in the crime after it has been committed. These investigations are conducted by crime scene investigation support. Preventive crime investigations involve raising awareness of the victims of crime by deterring criminals with effective anti-crime operations before the crime is committed. In addition, the concept of crime analysis is more important in preventive measures. This is because crime analysis is a concept in which the existing crime and criminal tendencies are detected and possible measures to be taken are also considered.

The use of crime in data mining through machine learning, identification, classification and clustering as patterns is also one of the important issues in various engineering and scientific disciplines such as biology, chemistry, physics, psychology, medicine and information for examining the evidence. This is because crime investigations are multidisciplinary (Hassani et al., 2016).

In several branches of science using criminal elements, some examples of usage areas are as follows:

· Examination of past handwritten letters and signatures with comparative examples taken from legal certificates and documentation,

· Comparison procedures of existing biological samples such as a nail, blood, and bone with existing databases within the scope of forensic biology,

· In scenes where serious accidents occurred, restructuring of the area with the aim of finding reasons for those accidents within the scope of traffic management and applications,

· Identification of a substance within the scope of forensic chemistry, especially its derivatives related to crime, such as drugs,

· Examination of the samples subject to comparison with past voice and video recordings within the scope of forensic audio and video examinations (Quick & Choo, 2016),

· Within the scope of forensic accounting, unusual money transfers or money transfers that are compatible with the amount subject to crime,

· Investigation of the connection of transactions made on a computer, internet, social media, mobile phone, etc. with crimes committed within the scope of computer forensics (Quick & Choo, 2016),

· In crime prevention activities to be carried out within the scope of criminology such as crime analysis, victim and suspect profiling, regions where crime is intense, etc. (Chan & Moses, 2017).

As can be seen from the above examples of usage areas, big data and data mining concepts are part of a multidisciplinary area. Their use is increasing with each passing day, yet not at the desired level. For this reason, some examples of usage areas of data mining and machine learning are given under the following main headings.

4.1. Corruption and fraud on bank accounts

The concept of corruption, counterfeiting and fraud is one of the crime concepts that is difficult to prevent by law enforcement. It is very difficult to identify the crime in its beginning stage since every emerging counterfeiting kind has a tendency to deceive people with new techniques. In addition, the concepts of corruption and fraud are becoming more and more complex with the developments in technology (Chau, Pandit, & Faloutsos, 2006). Together with unjust suffering, confidence in the state, and especially in state-affiliated security units, is being compromised in society and people are negatively affected.

With the development of technology, it is a known fact today that tangible money has turned into virtual money because of the ease of use of banks and even money statements in real amounts have been created using virtual transactions. In the money transfer data in question, problems are faced in most countries around the world only when they are in large amounts, banks share the necessary information with the relevant judicial authorities, so in fact data mining is being used (Odia & Akpata, 2020). However, notifying judicial authorities of suspicious money transfers, which are repeated frequently, even in small quantities, will facilitate the detection of illegal transactions from the very beginning (Agu et al., 2019).

Since the bank transactions we make daily happen quite frequently and in different amounts. While it is not easy to identify the elements that may be criminal in big data stacks, it is possible to determine the values that can be called suspicious by appropriate algorithmic analyses. For this reason, with machine learning works conducted within the scope of data mining, regular money transfers to an account from multiple accounts or to multiple shell accounts from an account can be detected and examining these transactions as suspicious transactions enables faster detection of crime. If a crime is detected in suspicious transactions, forensic law enforcement units will be involved, if no crime is detected, a preventive law enforcement action will be carried out, which is important in the prevention of crime. In addition, the comparison of the statistical data of customers in the bank with the usual behavioural patterns of the customers will help to identify any new suspicious activity. In this context, successful machine learning and deep learning models are available in the literature, especially in credit card and bank accounts (Adewumi & Akinyelu, 2017; Perols, 2011; Roy et al., 2018).

4.2. Using cell phone knowledge and base data

With the use of mobile phones worldwide being 68% of the population, especially in developed countries, it is obvious that the ratio is quite high (Wearesocials & Hootsuide, 2018). We can easily say that mobile phones are a part of our lives, as statistical data also reveals that they common. Considering the fact that the average mobile phone usage time in Europe is 251 minutes per month, it can easily be said that there is a high ratio of mobile phone usage (Ajans Press, 2017).

Our mobile phone use provides a lot of information such as with whom and how often we communicate in daily life, and on which geographical coordinate the phone receives the base station of the relevant GSM (Global System for Mobile Communications) company. In addition, it is possible to make frequent calls during the day and each of them creates data stacks at different time intervals. Such data stacks are called HTS (Historical Traffic Search) (In some sources it is referred to as "Call Data Record" (CDR)) (Steenbruggen et al., 2015) and they provide the knowledge on calls made by people on their phones including knowledge such as the caller, the called one, the call time, the call duration, the call location, and the received base stations (Boyd & Crawford, 2012). So, it becomes possible to find the approximate position of the person at the time of the crime and to establish the relationship between the suspect and the crime scene starting out from these meaningless data stacks. The I2 package program developed by IBM for this analysis is widely used for analysis purposes (Li, et al., 2006; Tassone, et al., 2017).

When the criminal investigations are examined, the suspects are usually people around the victim and related to the victim (Tilley & Sidebottom, 2017). Generally, there is a relationship between the suspect and the victim, as a crime is not committed without reason. Mental illnesses such as psychopathy are an exception to this relationship. One of the findings that reveal this relationship is the knowledge of past interviews, and the traffic of the last call before the event, especially when the victim is harmed, can often shed light on the event. In this context, there are successful models in the literature, especially regarding the use of mobile phones (He et al., 2020; Traunmueller et al., 2014).

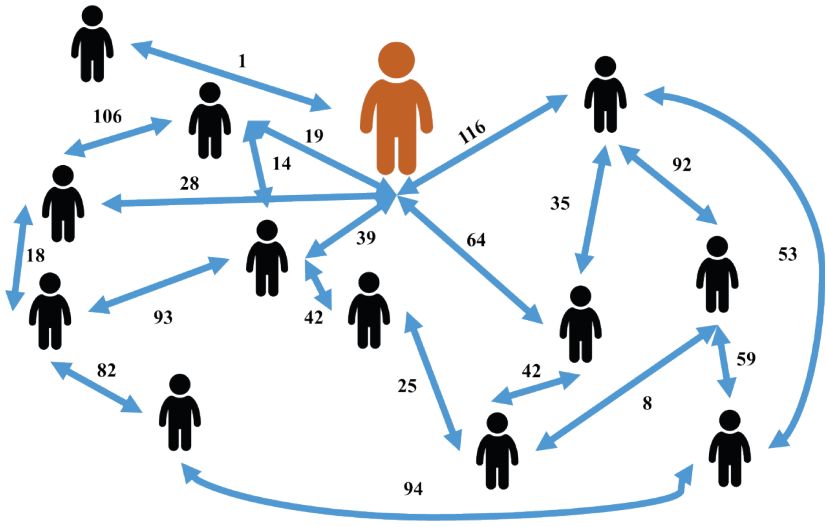


Figure 3. Example of HTS recording

As indicated in figure 3, HTS analysis is very important in regard to unfolding the relationship dimension, especially in organized crime organizations. For example, while one may assert that he/she is called by mistake in a one-time call of 10-seconds by the number x within a two-month period, it will be inconsistent to assert that he/she is also called by mistake for a total of 590 hours in 47 calls made by the number y.

4.3. Findings such as fingerprints and DNA in crime scenes

Some findings detected at the crime scene are quite important for particular identification, and fingerprints and DNA samples have a separate importance because of the fact that they provide detection so as not to leave room for suspicion (Bostanci, 2015; Wilson et al., 2010). Fingerprints start to form in humans when they are still a fetus. Fingerprints are used to identify people because each person's fingerprints are unique. In other words, no two people have exactly the same fingerprints so they are used as an identification purposes.

Whike the main papillary lines remain the same, injuries leave marks on the fingertip, but the main characteristic structure is always the same (Bhuyan, et al., 2010). All the fingerprint data in question constitutes the data stack. In forensic cases, the identities of the suspects are detected by making use of fingerprint findings taken from the crime scene where possible.

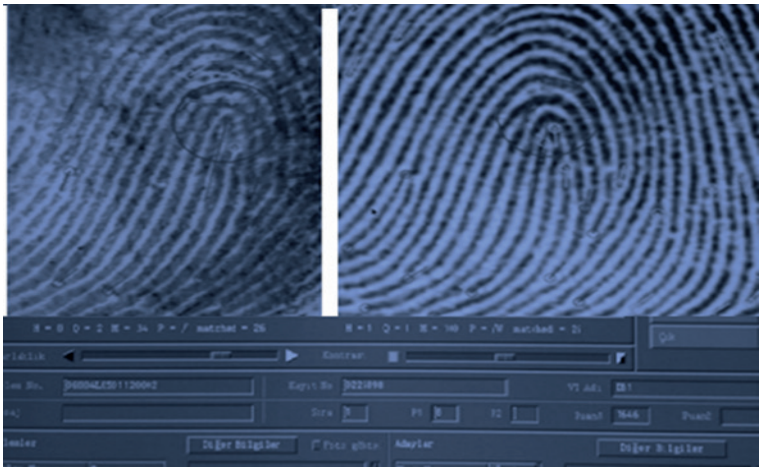


Figure 4. AFIS (Automatic Fingerprint Identification System) interrogation screen display

Figure 4 shows a comparison of fingerprints in the AFIS (Automated Fingerprint Identification System) database (Commission, 2017). The above-mentioned system scans the new fingerprint loaded into its memory and transfers it to its own database, and here it presents to experts the most similar fingerprint samples from the data pool by machine learning algorithms (Win et al., 2020). In this context, there are successful machine learning and deep learning models in the literature, especially in fingerprint recognition and classification (Uliyan et al., 2020; Wani et al., 2020; Pandya et al., 2018).

By comparing the DNA with the database, as in the example of fingerprints, if there is similar data in the data pool containing previously taken DNA samples, it is sent to experts for review (Xu, 2020). The only disadvantage of DNA is that DNA samples are identical in monozygotic twins. There are successful machine learning and deep learning models in the literature, especially in DNA (Aledhari et al., 2018; Lau & Fung, 2020).

4.4. Rulemaking from crime statistics and estimation of future crime tendency

Analysis of crime statistical information is a practice that involves revealing criminal tendencies by identifying existing crimes and also taking necessary precautions against such crimes (Lei, 2019). Because crime is a learned behaviour in general, and also "the past is the harbinger of the future" for future crime prediction. (Wang, et al., 2016).

Detecting criminal behaviour in the fight against crime is the most important topic of criminology. Criminal behaviour usually emerges as a result of different criminal motivations. The senses are the points of contact that connect individuals to the social world, and many psychological behaviours such as selfish behaviour and violence emerge together with the social world. The identification and categorization of these behaviours imply an understanding of the emotional state of the suspect himself/herself and of the society which includes him/her (Umair, et al., 2015). For this reason, crime statistics are important, and models which include a range of dynamic micro or macro areas that predict crime behaviour with hypothetical and experimental crime models have been developed in order to understand crime (Snaphaan & Hardyns, 2019; Bulgakova et al., 2019). However, the main problem in crime data is that it is in the form of a complex and large data stack before analysis. By ordinary statistical analysis methods, it is not always easy to obtain knowledge in this chaos also with the addition of dependent and independent variables that are in a specific position and interaction with others. For this reason, retrieval of meaningful knowledge from the said data stack is possible by the analysis of data. Use of data mining techniques gives the ability to proactively take action against criminal activities and potential security risks (Feng et al., 2019).

Many studies in criminology and sociology, regardless of the size of the specific analysis they define, provide a significant amount of knowledge on criminal density at levels such as micro and macro geographical location, criminal structure, etc. With the analysis of crime data in question:

Knowledge is available in many subjects such as types of crime mainly committed,
During which time crimes are mainly committed,
Where the crime hotspots are located,
Suspect profiling,
Victim profiling,
Whether there is a relationship or correlation between the crime type and other variables,
Estimation of the proportions and frequency of crime rates in the future.

Crime data are a relatively large volume of data. For example, the number of files received by the Offices of Chief Prosecutors in the Republic of Turkey (Population:83,154,997) in 2019 was recorded as 9,342,676 (transferred from last year, the total number of files received during the year) (TÜİK, 2020; Turkish Ministry

of Justice, 2020). When this number is examined, data mining techniques for the analysis of crime data, based especially on clustering, are widely used.

This sort of reasonable suspicion classification is only possible with the kind of prediction and rulemaking activity that is available through data mining. Such activity will provide the data explaining the crime to be revealed. In every crime, the determination of the crime profile with the objective modes of action to be put forward by the data will be the most important concept in fighting crime (Yoo, 2019). Studies in successful machine learning and deep learning are available in the literature, especially in crime prediction modelling (Berk, 2017; Mittal et al., 2018; Wheeler & Steenbeek,2020).

4.5 Analysis of the data subject to the crime on the internet and social media

In the information age, the concepts of time and space have become minimized with the emergence and development of internet, and major changes have occurred in many fields, especially in the field of communication. With these changes, people's leisure activities and attitudes changed, new communication methods were discovered, and many businesses, mainly journalism, education, banking, and trade services, started to operate in the virtual environment. The majority of recent works show that use of the internet has become a daily routine and that individuals spend a lot of time on the internet in order to pursue many daily activities. By means of increasing the number of technological devices that they own, individuals have reached the point where they can access the internet at any time, especially with devices such as portable laptops, tablets, and mobile phones, whether in the private space or the public space.

The Internet and social networks have transformed the way individuals communicate and socialize. Therefore, the virtual environments have become popular platforms for communication. In the world where people are increasingly using the internet and social networks, the amount of data that is in use is increasing day by day (Agrahari & Rao, 2017; Ateş et al., 2020).

With the emergence of the internet and social networking sites as a social platform, it has been observed that an increase in crimes committed over the internet and via social media have increased, especially in recent years, particularly crimes such as child pornography, cyberbullying, harassment, insult, internet fraud and cyber terrorism (Heickerö, 2014). Sometimes social networks can be used to organize human actions against the current government, as seen in the Arab spring and Gezi protests (Turkey). In these events, the organization of protests was achieved through Twitter and the

governments put a ban on Twitter. It is claimed that a system developed to classify user accounts as either being a supporter or non-supporter, achieved 90% accuracy while analysing Twitter accounts related to the Gezi protests (Yavanoglu, et al., 2013).

Shares made via the internet and social media also have an effect on clarifying crimes. In a study conducted in the United States, above 80% of law enforcement units actively use social networks (Guenther, 2012). Although there is no clear statistical study related to this issue, the internet, and particularly social networks on the internet, are actively being used for all kinds of forensic crime as well as for security investigations in most countries worldwide. For example, in a forensic case that is considered suicide, the normality of the person's posts will raise the suspicion that this may be murder that has been made to look like suicide, while in terrorism crimes, a person's posts praising a terrorist organization and sharing thoughts and pictures of the leader of the organization will also raise suspicions that this person may be related to the organization. Apart from these, various applications and programs on the internet, especially contact-based search engine scans and contact sharing applications (such as getcontact and truecaller) can support data on crime.

As well as data on the internet and on social networks (typically called "Social Networking Services" (SNS)) being used at a person-based level, it is also actively being used in preventive and informative tasks in the general framework (Arshad et al., 2019). 60-80% of intelligence over the world is obtained from open-source, and therefore analysis of the open-source data is becoming increasingly important (Power, 2016; Chen et al., 2014). Since the terrorist attacks of September 11, 2001, concerns about national security have significantly increased. Various intelligence units are actively collecting and analysing knowledge on various issues, especially the activities of terrorists (Power, 2016). If a crime is detected, the related units can be contacted again about judicial proceedings. Particularly through social networking tools that allow searching for keywords, general sharing of relevant topical issues can also be examined (Ayre & Craner, 2017). In addition, successful machine learning and deep learning models are available in the literature, especially in internet and social media studies (Ch et al., 2020; Williams et al., 2020; Ristea et al., 2020; Muneer & Fati, 2020).

4.6. Use of biometric properties in the security area

Biometrics authentication is typically used to measure the physical and behavioural properties of people. It is used in various security areas, particularly in identity validation,

by enabling the said measurable values to be distinguished through automation systems (Stewart, 2019). The most important property of the biometrics concept is that they are only used by themselves without using any object or data in identifications.

Biometric systems are mainly studied under two groups: physical (passive) and behavioural (active) (Tiwari, et al., 2015). In physical biometrics, voice, face, hand geometry, fingerprint, iris, and retina are used, while characteristics such as writing style, signature, walking patterns and lip movements during speech are used in behavioural biometrics. Essentially, physical biometrics is based on fixed physical properties of an individual that enable him/her to be distinguished from other people (Martinovic, et al., 2017). Behavioural biometrics is based on behaviours that are carried out by people different from each other in line with a specific purpose and at a certain time. The reliability of these biometrics is very high in terms of security because they are not transferable data as is the case with other validation methods.

Areas of daily life where biometric systems can be used are primarily in passports or imaging systems, at border control, in video surveillance, criminal identification, access control, computer logins, user verification in smartphones, crowd scanning, e-commerce, electronic banking, computer forensics, and ID cards.

Biometric verification is a very reliable method because it uses data mining based on supervised learning of patterns. Moreover, the design and simulation of such systems also becomes much simpler using artificial nervous systems and signal processing techniques. Considering its positive aspects of greater identity validation and security, it is expected that the frequency of use of biometrics will increase in the near future. Successful machine learning and deep learning models are available in the literature, especially on the use of biometrics (Adamović et al., 2020; Arora et al., 2020; Pandey et al., (2017); Sundararajan & Woodard, 2018).

5. Discussion

In operations carried out on crime data, the priority is that law enforcement officers act proactively and prevent crime before it is committed. When crime is prevented, there will be no victims. However, it is not easy to intervene at the point when the criminal activities occur . In addition, since no crime has been committed, the guilty person only receives a small punishment for the attempted act. Victimization should be avoided, even if the suspects will receive less punishment. In this context, in recent years, highly successful crime prevention activities have been carried out, especially

on digital platforms, using deep learning algorithms acting on the basis of machine learning.

The repetition of criminal behaviour committed by the same people in the same places is a well-known subject in the theory of hot spots and routine activities (Song et al., 2019; Yao et al., 2020). As a result of analysing historical crime data with data mining techniques, these repetitions can be detected and successful crime prevention policies can be established.

Once a crime has been committed law enforcement officers intervene. At this stage, the most important issue is that every contact will leave a trace, as stated in Locard's change principle (Mistek et al., 2018). The traces in question are silent witnesses of the crime. The crime scene can be a physical space or a virtual space. Based on the crime scene, a relationship is attempted to be established between the suspect, the victim, the incident, and the scene (Bode, 2019).

Due to the development of technology in recent years, the processing of big data and the rapid results based on machine learning have made a great contribution in solving criminal cases, far greater than that achieved by human power (McCue, 2014). In operations carried out on crime data, making meaningful inferences constitutes evidence. Evidence is the fundamental basis of judicial authorities in the trial process. Decisions made on the basis of evidence will leave no doubt in anybody's mind and will increase society's trust in the judicial authorities.

6. Conclusion

Crime is one of the problems that society has faced throughout human history. For this reason, many areas of science have put forward various methods and tactics to fight existing crimes. The concept of crime has been transformed throughout human history, along with each new method for fighting crime, but in its essence, it has remained as a phenomenon that causes harm to the other party. For this reason, every society carries out operations which assist in the determination of the criminal profile responsible for the criminal acts, the detection of the causes of crime, the prevention of them, the fight against them, the exposure of particular crimes and the places where they are committed.

With the development of technology, we have a lot of data stack that we cannot actively utilize in many subjects, and we experience difficulties in the classification of this ever-increasing amount of data. Crime data related to criminology and criminality are

the main ones. For this reason, data mining process is clearly very important, because what matters is the retrieval of meaningful knowledge. In addition to data mining, along with the development of concepts such as artificial intelligence, machine learning and advances in technology, law enforcement units have increased their work in this area. The digital data obtained from these areas are combined with the data obtained from the physical world, allowing the judicial authorities to make some determinations by developing a hypothesis subject to crime.

The aim of this article was to provide an account of data mining applications in the areas of criminology and criminalistics. The examples given in the application section are only some of the main application areas and it is not possible to limit the topic of crime to these examples. The authors expect and predict that these methods will be used widely in both a preventive and reactive sense in the fight against crime.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.
Finansal Destek: Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

References/Kaynakça

Abdullah, N., Ismail, S. A., Sophiayati, S., & Sam, S. M. (2015). Data quality in big data: a review. *International Journal of Advances in Soft Computing & Its Applications*, *7*(3).

Adamović, S., Miškovic, V., Maček, N., Milosavljević, M., Šarac, M., Saračević, M., & Gnjatović, M. (2020). An efficient novel approach for iris recognition based on stylometric features and machine learning techniques. *Future Generation Computer Systems*, *107*, 144-157.

Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, *8*(2), 937-953.

Aggarwal, C. C. (2018). *Machine learning for text*. Cham: Springer International Publishing.

Agrahari, A., & Rao, D. (2017). A review paper on Big Data: technologies, tools and trends. *Int Res J Eng Technol*, *4*(10), 640-649.

Agu, S. C., Ajah, I., & Ibe, W. E. (2019). Impact of Human Character and Information System on Corruption Risk in Nigeria. *International Journal of Scientific Research and Engineering Development, 2(4), 481-485.*

Ahmed, A. (2020). "From Data to Wisdom" Using Machine Learning Capabilities in Accounting and Finance Professionals. *Talent Development & Excellence*, *12*.

Ajans Press, 2017. [Online]. Available: https://www.cnnturk.com/bilim-teknoloji/ turkiye-cep-telefonuyla-konusmada-avrupa-birincisi?page=1 (accessed 5.12.18).

Akdemir, N., & Lawless, C. J. (2020). Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: a lifestyle routine activities approach. Internet Research, 30(6), 1665-1687.

Aledhari, M., Di Pierro, M., Hefeida, M., & Saeed, F. (2018). A deep learning-based data minimization algorithm for fast and secure transfer of big genomic datasets. *IEEE Transactions on Big Data*.

Arora, S., Bhatia, M. P. S., & Kukreja, H. (2020, February). A Multimodal Biometric System for Secure User Identification Based on Deep Learning. In *International Congress on Information and Communication Technology* (pp. 95-103). Springer, Singapore.

Arshad, H., Jantan, A., & Omolara, E. (2019). Evidence collection and forensics on social networks: Research challenges and directions. *Digital Investigation*, *28*, 126-138.

Ateş, E.C., Bostanci, E., & Guzel, M. S. (2020). Security Evaluation of Industry 4.0: Understanding Industry 4.0 on the Basis of Crime, Big Data, Internet Of Thing (IoT) and Cyber Physical Systems. *Güvenlik Bilimleri Dergisi*, (International Security Congress Special Issue), 29-50.

Ayre, L. B., & Craner, J. (2017). Open data: What it is and why you should care. *Public Library Quarterly*, *36*(2), 173-184.

Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. *International journal of engineering research & technology (IJERT)*, *1*(6), 1-6.

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, *13*(2), 193-216.

Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.

Bhuyan, M. H., Saharia, S., & Bhattacharyya, D. K. (2012). An effective method for fingerprint classification. *arXiv preprint arXiv:1211.4658*.

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689-8692.

Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., & Klusemann, B. (2019). A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials*, *6*, 110.

Bode, J. (2019, June). Every Contact Leaves a Trace: A Literary Reality of Locard's Exchange Principle. In *Outside the Box: A Multi-Lingual Forum* (p. 18).

Bostanci, E. (2015). 3D reconstruction of crime scenes and design considerations for an interactive investigation tool. *arXiv preprint arXiv:1512.03156*.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679.

Bulgakova, E., Bulgakov, V., Trushchenkov, I., Vasilev, D., & Kravets, E. (2019). Big data in investigating and preventing crimes. In *Big Data-driven World: Legislation Issues and Control Technologies* (pp. 61-69). Springer, Cham.

Campbell, C., & Ying, Y. (2011). Learning with support vector machines. *Synthesis lectures on artificial intelligence and machine learning*, *5*(1), 1-95.

Ch, R., Gadekallu, T. R., Abidi, M. H., & Al-Ahmari, A. (2020). Computational System to Classify Cyber Crime Offenses Using Machine Learning. *Sustainability*, *12*(10), 4087.

Chan, J., & Bennett Moses, L. (2017). Making sense of big data for security. *The British journal of criminology*, *57*(2), 299-319.

Chau, D. H., Pandit, S., & Faloutsos, C. (2006, September). Detecting fraudulent personalities in networks of online auctioneers. In *European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 103-114). Springer, Berlin, Heidelberg.

Chen, M., Mao, S., & Liu, Y. (2014). *Big data: A survey.* Mobile networks and applications, 19(2), 171-209.

Clarke, C. (2006). Proactive policing: Standing on the shoulders of community□based policing. *Police Practice and Research*, *7*(1), 3-17.

Commission, (2017). Kriminalistik. Gendarmerie and Coast Guard Academy, Ankara.

Cooper, P. (2017). Data, information, knowledge and wisdom. *Anaesthesia & Intensive Care Medicine*, *18*(1), 55-56.

Dey, A. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, 7(3), 1174-1179.

Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y., & Liu, Q. (2019). Big data analytics and mining for effective visualization and trends forecasting of crime data. *IEEE Access*, *7*, 106111-106123.

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *Special Section On Data-Driven Monitoring, Fault Diagnosis and Control Of Cyber-Physical Systems*, *5*, 20590-20616.

Ghorbani, R., & Ghousi, R. (2019). Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *International Journal of Data and Network Science, 3(2), 47-70.*

Guenther, A.J. (2012). Role of Social Media in Law Enforcement Significant and Growing [Online]. Available: http://www.lexisnexis.com/en-us/about-us/media/press-release.page?id =1342623085481181

Gupta, M. K., & Chandra, P. (2019, March). A comparative study of clustering algorithms. In *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 801-805). IEEE.

Hand, D. J., & Adams, N. M. (2014). Data Mining. Wiley StatsRef: Statistics Reference Online, 1-7.

Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *9*(3), 139-154.

He, L., Páez, A., Jiao, J., An, P., Lu, C., Mao, W., & Long, D. (2020). Ambient Population and Larceny-Theft: A Spatial Analysis Using Mobile Phone Data. *ISPRS International Journal of Geo-Information*, *9*(6), 342.

Heickerö, R. (2014). Cyber terrorism: Electronic jihad. Strategic Analysis, 38(4), 554-565.

Hey, J. (2004). The data, information, knowledge, wisdom chain: the metaphorical link. *Intergovernmental Oceanographic Commission*, *26*, 1-18.

Jackson, J. (2002). Data mining; a conceptual overview. *Communications of the Association for Information Systems, 8(1), 19.*

Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.

Khare, A. R., & Shrivasta, P. (2018). Data mining for the internet of things. In *Exploring the Convergence of Big Data and the Internet of Things* (pp. 181-191). IGI Global.

Koyuncugil, A. S., & Özgülbaş, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *İnternational Journal Of Informatics Technologies*, *2*(2).

Kumar, R., & Nagpal, B. (2019). Analysis and prediction of crime patterns using big data. International Journal of Information Technology, 11(4), 799-805.

Lau, P. Y., & Fung, W. K. (2020). Evaluation of marker selection methods and statistical models for chronological age prediction based on DNA methylation. *Legal Medicine*, *47*, 101744.

Lei, C. (2019). Legal Control over Big Data Criminal Investigation. *Social Sciences in China*, *40*(3), 189-204.

Li, X., Liu, B., & Philip, S. Y. (2006, September). Discovering overlapping communities of named entities. *In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 593-600).* Springer, Berlin, Heidelberg.

Martinovic, I., Rasmussen, K., Roeschlin, M., & Tsudik, G. (2017). Authentication using pulse-response biometrics. *Communications of the ACM*, *60*(2), 108-115.

Mcclendon, L., & Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ), 2(1), 1-12.*

McCue, C. (2014). *Data mining and predictive analysis: Intelligence gathering and crime analysis*. Butterworth-Heinemann.

Mesgarpour, M., & Dickinson, I. (2014). Enhancing the value of commercial vehicle telematics data through analytics and optimisation techniques. *Archives of Transport System Telematics, 7.*

Mistek, E., Fikiet, M. A., Khandasammy, S. R., & Lednev, I. K. (2018). *Toward locard's exchange principle: recent developments in forensic trace evidence analysis.* Analytical chemistry, 91(1), 637-654.

Mittal, M., Goyal, L. M., Hemanth, D. J., & Sethi, J. K. (2019). Clustering approaches for high□dimensional databases: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1300.

Mittal, M., Goyal, L. M., Sethi, J. K., & Hemanth, D. J. (2018). Monitoring the impact of economic crisis on crime in India using machine learning. *Computational Economics*, *53*(4), 1467-1485.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C. (2013). A survey of multiobjective evolutionary algorithms for data mining: Part I. *IEEE Transactions on Evolutionary Computation*, *18*(1), 4-19.

Muneer, A., & Fati, S. M. (2020). A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter. *Future Internet*, *12*(11), 187.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, *36*(2), 2592-2602.

Odia, J. O., & Akpata, O. T. (2020). Role of Data Science and Data Analytics in Forensic Accounting and Fraud Detection. In *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics* (pp. 203-227). IGI Global.

Olson, D. L., & Lauhoff, G. (2019). Descriptive data mining. In *Descriptive Data Mining* (pp. 129-130). Springer, Singapore.

Pandey R.K., Zhou Y., Kota B.U., Govindaraju V. (2017) Learning Representations for Cryptographic Hash Based Face Template Protection. In: Bhanu B., Kumar A. (eds) *Deep Learning for Biometrics. Advances in Computer Vision and Pattern Recognition.* Springer, Cham. https://doi.org/10.1007/978-3-319-61657-5_11

Pandya, B., Cosma, G., Alani, A. A., Taherkhani, A., Bharadi, V., & McGinnity, T. M. (2018, May). Fingerprint classification using a deep convolutional neural network. In *2018 4th International Conference on Information Management (ICIM)* (pp. 86-91). IEEE.

Pauleen, D. J., Rooney, D., & Intezari, A. (2017). Big data, little wisdom: trouble brewing? Ethical implications for the information systems discipline. *Social Epistemology*, *31*(4), 400-416.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, *30*(2), 19-50.

Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision systems, 25(sup1), 578-588.*

Quick, D., & Choo, K. K. R. (2016). Big forensic data reduction: digital forensic images and electronic evidence. *Cluster Computing, 19(2), 723-740.*

Ristea, A., Al Boni, M., Resch, B., Gerber, M. S., & Leitner, M. (2020). Spatial crime distribution and prediction for sporting events using social media. *International Journal of Geographical Information Science, 1-32.*

Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018, April). Deep learning detecting fraud in credit card transactions. In *2018 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 129-134). IEEE.

Rutkowski, L., Jaworski, M., & Duda, P. (2020). *Stream data mining: algorithms and their probabilistic properties*. Cham: Springer.

Shao, L., Duan, Y., Sun, X., Gao, H., Zhu, D., & Miao, W. (2017, July). Answering Who/When, What, How, Why through Constructing Data Graph, Information Graph, Knowledge Graph and Wisdom Graph. In *SEKE* (pp. 1-6).

Snaphaan, T., & Hardyns, W. (2019). Environmental criminology in the big data era. *European Journal of Criminology*, 1477370819877753.

Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., & Liao, W. (2019). Crime feeds on legal activities: Daily mobility flows help to explain thieves' target location choices. *Journal of Quantitative Criminology*, *35*(4), 831-854.

Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE), 2(02), 250-255.*

Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities?. *Telecommunications Policy*, *39*(3-4), 335-346.

Stewart, L. (2019). Big Data Discrimination: Maintaining Protection of Individual Privacy without Disincentivizing Businesses' Use of Biometric Data to Enhance Security. *BCL Rev.*, *60*, 349.

Sundararajan, K., & Woodard, D.L. (2018). Deep Learning for Biometrics: A Survey. *ACM Comput. Surv. 51(3), DOI:https://doi.org/10.1145/3190618.*

Tassone, C., Martini, B., & Choo, K. K. (2017). Forensic visualization: survey and future research directions. *In Contemporary Digital Forensic Investigations of Cloud and Mobile Applications (pp. 163-184). Syngress.*

Tilley, N., & Sidebottom, A. (2017). *Handbook of crime prevention and community safety.* Routledge.

Tirgari, V. (2012). Information technology policies and procedures against unstructured data: A phenomenological study of information technology professionals. *Journal of Management Information and Decision Sciences, 15(2), 87.*

Tiwari, S., Chourasia, J. N., & Chourasia, V. S. (2015). A review of advancements in biometric systems. *International Journal of Innovative Research in Advanced Engineering, 2(1), 187-204.*

Traunmueller, M., Quattrone, G., & Capra, L. (2014, November). Mining mobile phone data to investigate urban crime theories at scale. *In International Conference on Social Informatics (pp. 396-411). Springer, Cham.*

Turkish Ministry of Justice. (2020). Judicial Statistics 2019. [Online]. Available: https://adlisicil.adalet.gov.tr/Resimler/SayfaDokuman/1092020162733adalet_ist-2019.pdf (accessed 10.09.20).

Turing, A., (1950). Computing machinery and intelligence: Mind, 59, 433–460.

TÜİK (Turkish Statistical Institute). (2020). Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları, 2019. [Online]. Available: https://adlisicil.adalet.gov.tr/Resimler/SayfaDokuman /1092020162733adalet_ist-2019.pdf (accessed 10.09.20).

Umair, S., Muhammad, S., Amna, U., Aniqa, M., Abdul, B.S., Sheikh, K.R., (2015). Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining. *Res. J. Recent Sci. (pp. 106–114).*

Uliyan, D. M., Sadeghi, S., & Jalab, H. A. (2020). Anti-spoofing method for fingerprint recognition using patch based deep learning machine. Engineering Science and Technology, an International Journal, 23(2), 264-273.

Vaidhyanathan, S., & Bulock, C. (2014). Knowledge and dignity in the era of "big data". *The Serials Librarian, 66(1-4), 49-64.*

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165, 234-246.*

Wang, H., Kifer, D., Graif, C., & Li, Z. (2016, August). Crime rate inference with big data. *In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 635-644).*

Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2020). Supervised Deep Learning in Fingerprint Recognition. In Advances in Deep Learning (pp. 111-132). Springer, Singapore.

Wearesocials & Hootsuide, (2018). Digital in 2018: World's ınternet users pass the 4 billion mark. URL https://wearesocial.com/blog/2018/01/global-digital-report-2018 (accessed 16.02.20).

Wheeler, A. P., & Steenbeek, W. (2020). Mapping the risk terrain for crime using machine learning. *Journal of Quantitative Criminology*, 1-36.

White, M. (2012). Digital workplaces: Vision and reality. *Business information review*, *29*(4), 205-214.

Williams, G. J. (2009). Rattle: a data mining GUI for R. The R Journal, 1(2), 45-55.

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, *60*(1), 93-117.

Wilson, D. B., McClure, D., & Weisburd, D. (2010). Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions. *Journal of Contemporary Criminal Justice, 26(4), 458-469.*

Win, K. N., Li, K., Chen, J., Viger, P. F., & Li, K. (2020). Fingerprint classification and identification algorithms for criminal investigation: A survey. Future Generation Computer Systems, 110, 758-771.

Xu, H. (2020). Big data challenges in genomics. In Handbook of Statistics (Vol. 43, pp. 337-348). Elsevier.

Yao, S., Wei, M., Yan, L., Wang, C., Dong, X., Liu, F., & Xiong, Y. (2020, August). *Prediction of Crime Hotspots based on Spatial Factors of Random Forest.* In 2020 15th International Conference on Computer Science & Education (ICCSE) (pp. 811-815). IEEE.

Yavanoglu, U., Colak, M., Caglar, B., Cakir, S., Milletsever, O., & Sagiroglu, S. (2013, December). Intelligent approach for identifying political views over social networks. *In 2013 12th International Conference on Machine Learning and Applications (Vol. 2, pp. 281-287). IEEE*.

Yoo, J. S. (2019, December). Crime data warehousing and crime pattern discovery. In *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems* (pp. 1-6).

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology*, *58*(4), 479-493.