Utilization of Big Data in Criminal Case Disclosure by the Indonesian Police Investigators

Wira Prayatna¹, Alvi Syahrin², Saidin³ and Keizerina Devi Azwar⁴

Abstract

The benefit of using Big Data in uncovering criminal cases is its ability to trace the digital footprint of criminals. The purpose of this paper is to find out the use of big data in criminal cases by the Indonesian National Police. A digital footprint is a physical copy of photos, voice posts, and a criminal event that can be traced to reveal the perpetrator’s crime. Indonesian Police investigators do not need to spend a lot of time searching and verifying digital footprints because Big Data can be used to collect such information for analysis to find both the criminal events and the perpetrators. The use of big data in disclosing a crime has not been institutionalized in the investigation process carried out by the Indonesian Police investigators, due to the influence of the Negatief Wettelijk Bewijsleer-based proof system applicable in the proof system of the Criminal Procedure Code. The problem of this paper is how to uncover criminal cases by utilizing Big Data. A doctrinal approach was used to answer the problem. The results of the study indicate that the use of Big Data by the Indonesian Police investigators has not been followed by a regulation in the form of a criminal policy in the framework of evidence based on the Criminal Procedure Code. Digital evidence in the legal system in Indonesia has been regulated in the Law on Electronic Information and Transaction, but it has not been followed by an integrated coherence in the criminal proof system in the Criminal Procedure Code.

Keywords: Big Data, Crime Case Disclosure.

INTRODUCTION

Big data can be used as a means of disclosing crimes. It is possible because of the very rapid current development of Information and Communication Technology. The emergence of electronic engineering, computerization, artificial intelligence, biochemistry, architecture, material engineering science, etc. does not only provide a technical basis for human benefits but also opens opportunities for violations of individual rights through its ability to analyze and visualize the data associated with the individuals. In the Internet sector, there has been rapid progress with the introduction of the Internet of Things (IoT) which is very useful to support the development of Big Data technology and Computing Technology for various purposes including prevention, countermeasures, and disclosure of criminal cases. These findings can also serve as tools that can be used in crime prevention and enable the emergence of new crime prevention technologies and systems. i For example, the development of Algorithmic Regulation has become a central theme in the discussion of current regulatory policies in various countries. ii Regulating Big Data in Indonesia is a form of transplanting a foreign legal system. In addition, the use of Big Data leads to a clash of evidentiary legal norms and a vacuum in legal norms because the Law on Electronic Information and Transaction (EIT Law) in Indonesia only recognizes electronic evidence. Therefore, Big Data is only used as a supporting tool to uncover criminal cases. Transplantation of EIT Law in the form of electronic evidence is not followed by a coherent culture of law enforcement in Indonesia which has been familiar with the use of the Criminal Code and Criminal Procedure Code instruments. Enforcement of Indonesian criminal law in the framework of evidence is more oriented towards criminal responsibility based on evidence which, in the Criminal Procedure Code, is known as the influence of “legalistic positivism” on the civil law system. Transplantation of Big Data in the EIT Law is due to the occurrence of democratization in various existing

¹ Universitas Sumatera Utara, Medan, Indonesia, E-mail: pol.wiraprayatna@gmail.com
² Universitas Sumatera Utara, Medan, Indonesia
³ Universitas Sumatera Utara, Medan, Indonesia
⁴ Universitas Sumatera Utara, Medan, Indonesia
countries and the formation of countries in the process of decolonization of western countries. Institutional transplantations occur more frequently and are often enforced by their home countries.

The problem of this paper is formulated into the question “How is Big Data used to uncover criminal cases?” This study used a doctrinal approach (doctrinal research method) which analyzed both laws as it is written in the books and as it is decided by the judge through a judicial process. The data were secondary data derived from the cases taking place in the United States which were compared to the ones taking place in Indonesia. The data were analyzed using a descriptive qualitative method.

RESULTS AND DISCUSSION

Big Data Using Computing Technology and Internet of Things in Uncovering Criminal Cases

The attention to Big Data has been emerging since 2011, then interest in this field continues to grow rapidly. Unlike most studies in computer science, Big Data has received significant attention from various media. News about Big Data, its benefits, and possible violations of privacy due to its usage has been a topic in various media. In addition, since its emergence, the discussion about Big Data has been related to technical and socio-technical issues; nevertheless, the exact definition of Big Data remains unclear to date. Due to a large number of sources talking about Big Data, there is no uniform definition of Big Data which leads to contradictions and discrepancies. Microsoft provides a concise definition of Big Data as “a term that is increasingly being used to describe the process of applying serious computing power, which is the latest in machine learning and artificial intelligence for massive data and dozens of highly complex sets of information”. In its development, the term Big Data has gradually gained a universal form. However, it varies depending on different sources. Academia, industry, and media make their own definitions of Big Data which are certainly different from one another. In addition, all stakeholders also provide different, even contradictory, definitions of Big Data.

Dawn E. Holmes states:

“Big Data is now used to refer not just to the total amount of data generated and stored electronically, but also to specific datasets that are large in both size and complexity, with which new algorithmic techniques are required in order to extract useful information from them. These big datasets come from different sources, so let’s take a more detailed look at some of them and the data they generate”. [v]

Then, Janne Tarkoma argues that:

“Big Data denotes data of a very large size. Typically the management and manipulation of these large data sets is logistically challenging. In technology, the term can also be used to describe the branch of computing that involves such data. The use of different innovative techniques enables the commercialization of raw data (or primary data), which refers to the unprocessed form of big data collected from the source”. [vi]

Meanwhile, the definition of Big Data given by the leading ICT Consulting Firm, Gartner, is:

“Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization”. [vii]

The non-uniform definition confuses and hinders general discussion about Big Data. However, in general, the characteristics of Big Data are: 1) The volume (amount) is very large as the total size of the data is usually more than 1 terabyte; 2) The velocity (growth) of the data is very fast as the data increase in a very large amount in a short time; 3) The variety of data (various forms or formats of data) can be in the form of plain text files, database tables (e.g. MySQL, Excel files), or any other forms. Some parties add the next two “V” characteristics, namely Veracity (data accuracy and its processing results) and Value (the value generated by the data), but the above definition is sufficient to describe the shape of Big Data in general. [viii]

Big Data has special characteristics because Big Data functions with both structured and unstructured data. Structured data (e.g. spreadsheets, timesheets, etc.) are neatly organized, so they are relatively easier to manage.
to become valuable information. Unstructured data (e.g. email, video, etc.) are in the form of millions of emails, so they are relatively difficult to manage to become valuable information. In processing complex data, distributed computing environments and Massively Parallel Processing (MPP) architectures are often used. Their usage is more appropriate for data absorption and analysis with a parallel approach.

Authorities and regulators have defined Big Data as a collection of various factors, including the collection of data that exist in various places and come from various sources. Nowadays, that data storage costs tend to decrease and the capacity obtained is getting bigger allows more intensive activities to process data. Data collected from various sources generally have the following characteristics: proprietary data, open data, creative commons data, and public domain data.

Although the term Big Data is not yet known in Indonesian legislation, in general, the regulations regarding ‘data’ have been included in the EIT Law. Article 1.1 of EIT Law states that Electronic Information is one or a set of electronic data, including but not limited to writing, voice, image, map, design, photograph, Electronic Data Interchange (EDI), electronic mail, telegram, telex, telecopy, or the like, processed letters, signs, numbers, access codes, symbols or perforations that have meaning or can be understood by people who can understand them. The Indonesian Government Regulation Number 82 of 2012 includes the same definition of data. Article 1.6: states that electronic information is one or a set of electronic data, including but not limited to writing, voice, image, map, design, photograph, Electronic Data Interchange (EDI), electronic mail, telegram, telex, telecopy, or the like, processed letters, signs, numbers, access codes, symbols, or perforations that have meaning or can be understood by those who can understand them. Electronic data is a part of electronic information, regardless of the volume, velocity, and variety of the data. Thus, Big Data has been implicitly covered by both the law and the regulation. Moreover, these laws and regulations should be dynamic and should not stop being developed because data can change rapidly.

The value of Big Data will only emerge if it is used to assist decision-making in companies, government, and law enforcement offices. To enable Big Data-based decision making needs a process of converting a high volume of data that quickly moves, increases, changes, and varies into a set of structured and meaningful data.

Law Enforcement Officers who take advantage of Big Data in making decisions will make just laws. Law and justice are actually two interrelated elements that constitute “conditio sine quo non” for the other. Law is an external manifestation of justice, and justice is an authentic internal and the spiritual essence of the form of law. Therefore, the supremacy of law is the supremacy of justice, and vice versa, because both are commutative things.

The highest goal of the law is justice, i.e. to put something in accordance with its proportion. So, justice is both order and discipline. Everything must be directed to find a legal system that is most appropriate and following the principle of justice. Law is a just constitution, so if a constitution is not following the principle of justice, it cannot be said to be the law.

The whole process of extracting (retrieving) information from Big Data can be divided into five stages: 1) Acquisition and Recording Stage; 2) Extraction, Cleaning and Annotation Stage; (3) Integration Stage; 4) Aggregation and Representation, Modeling and Analysis Stage; and 5) Data Interpretation Stage. These five stages form two main sub-processes: data management and data analysis. Data management involves supporting processes and technologies of collecting and storing data as well as preparing and retrieving data for analysis. Data analysis, on the other hand, refers to the techniques used to analyze and acquire knowledge from Big Data. Big Data analysis can be seen as a sub-process in the whole process of extracting information from Big Data.

Big Data gathered from social media, e.g. Facebook, Twitter, LinkedIn, etc., can be used by processing it with business data from a company, such as customer data that have already been in the company’s database, product sales data, etc. Big Data is then processed to compile a marketing plan for the company’s product. Big Data is used by businessmen for various purposes. The classic examples of data use are Uber and its taxis, and Amazon and its bookstores, where Uber analyzes and uses large amounts of information to provide a service.
The utilization of Big Data is inseparable from the support of computing technology that is growing rapidly. Today’s computing technology is changing. Cloud technology (cloud) allows the storage of applications and documents into the cloud which consists of thousands of computers and servers that are linked together and can be accessed via the internet. All of the activities on the computer are done via web-based cloud technology, not desktop-based technology.

Desktop-based documents and applications are moved to the cloud. People will no longer be tied to a computer located in their offices because the data can be stored on the web and can be accessed from anywhere in the world. Although early cloud computing used to be very doubtful, it is now being extensively used by providers (general providers) such as: Gmail, Hotmail, and Apple me Mobile. Several large companies, e.g. Google, Amazon, Microsoft, and sun system offer cloud computing with a growing variety of services. [xxv] Cloud computing services are increasingly developing making them more affordable in terms of ease of use and terms of cost; even many service providers nowadays do not charge the users for a certain amount of storage capacity. The cloud sensor system reduces losses from fire, water leakage, as well as theft, and accidents at home. By combining sensors, cameras, and careful analytics, future IoT systems can detect when people are in danger and immediately alert the fire brigade, police, or emergency services to take swift action. [xxv] IoT-Based Criminals Detection and Tracking is an IoT-based mobile collaboration system that is used to detect and track tailored to individual customers. Amazon analyzes millions of data to be able to recommend to its consumers the best books to read. [xviii]

Cloud computing can be used as a network to perform storage, distribution, expenditure, and other computing processes as needed. Cloud computing enables its users to accurately obtain the computing resources needed by and for certain computing through the distribution network infrastructure. Cloud computing takes three forms: public cloud, private cloud, and mixed cloud. Generally, the form first built on the computer is a private cloud, and it is gradually extended to the public cloud when the technology matures. [xxi]

Cloud computing services consist of three types: IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service). IaaS provides the infrastructure services needed. The technology used is virtualization technology which changes the approach of information platform construction and the mode of information service provision. IaaS offers a technical basis for reusing resources and increasing flexibility, efficiency, and accuracy of resource allocation. PaaS provides software platform services consisting of software testing, deployment, operation environment, and application services. Besides, PaaS also provides an effective sharing mechanism of computation capability for the whole open circle of application development. SaaS provides users with several software applications that can be run in a cloud platform. [xxii]

Internet of things (IoT) is a physical device working system with the help of the internet. IoT is a physical device that is completed with sensors, electronics, software, and network connectivity, enabling it to collect and exchange data. [xxiii] The IoT has the potential to fundamentally change the way people interact with the environment. The IoT’s ability to electronically monitor and manage objects in the physical world enables people to optimize performance process systems, save time, improve quality of life, and efficiently make data-driven decisions in almost all human activities. Today’s IoT has made human life easier because just with a smartphone, people can access the IoT and control various devices remotely. The emergence of such a smart device is the need for people to connect to the Internet more easily from various places. Today, with smart devices, people can do more than just connect to the Internet and simply monitor their heart rate or sleeping time. People use smart devices to connect with other people, pay bills, shop, rent a car or house, book theater tickets, or airplane tickets, connect people to their bank accounts, and much more. The use of IoT has become much wider and has shifted to smart-home, smart-car, smart-farming, smart-city, and others. [xxiv]

The IoT sensor system reduces losses from fire, water leakage, as well as theft, and accidents at home. By combining sensors, cameras, and careful analytics, future IoT systems can detect when people are in danger and immediately alert the fire brigade, police, or emergency services to take swift action. [xxv] IoT-Based Criminals Detection and Tracking is an IoT-based mobile collaboration system that is used to detect and track
Utilization of Big Data in Exposing Crimes

In general, the data collected are those that have or are subject to copyright, database rights, moral rights, trademark rights, and rights to trade secrets or confidential information. The use of personal data can lead to legal action cases filed by individuals whose data are used illegally. In practice, the use of data is often without the permission of the data owner, especially concerning personal data, because consent to data use is usually included only in the terms and conditions when someone signs up for a particular service provider, such as Facebook, Twitter, Yahoo (e-mail), and so forth. There are legal problems with the use of Big Data. Law enforcers must pay attention to the security of personal data, legal protection of data, and use of data by other parties as well as by social media providers themselves.

One of the examples is the case of Spokeo v. Robins. In this case, Spokeo, Inc. operates a website that provides information about an individual such as contact data, marital status, age, occupation, and certain types of economic information. Thomas Robins sued Spokeo and claimed that Spokeo deliberately violated the Fair Credit Reporting Act (FCRA) by publishing false information about him on the website. However, Robins was unable to specify the actual or future harm incurred, so the District Court granted Spokeo’s motion to dismiss the case, due to lack of jurisdiction over the subject matter and Robins’ lack of stance under Article III of the Constitution. Robins then filed an amendment to the lawsuit, in which he alleged that he had obviously suffered a real loss to his job prospects because the website “deliberately” claimed he was a huge fortune. The district court initially rejected Robins’ lawsuit, but later reconsidered the decision and ultimately rejected the suit because it did not provide details of actual damages. Robins appealed and stated that the District Court could not reconsider the previous decision and that he had filed enough real loss to meet the requirements of the Constitution. The U.S. Court of Appeals for The Ninth Circuit dismissed the District Court's decision and stated that “even though the District Court may reconsider its decision, the allegation of violation of legal rights is sufficiently detrimental and qualifies to file a lawsuit”. [xxvii]

Another example is the joined cases C-293/12 and C 594/12. On April 8, 2014, the Grand Chamber of the European Court of Justice (ECJ) stated that the “Directive 2006/24/EC” regarding the storage of data generated in connection with the provision of publicly available electronic communication services or public communication networks (the so-called “Data Retention Directive” or DRD) is not legally valid. The legal framework established by DRD - which largely consists of obligations imposed after a service provider systematically stores all traffic and location data of all electronic communication users (Internet e-mail, telephone service) for a minimum period of 6 to a maximum of 24 months has been proven to violate the basic rights to privacy and protection of personal data. [xxviii] The court considers there is an interference into the private domain of an individual in collecting data related to electronic communications. If needed to combat international terrorism or serious crimes, such interferences can be justified. The decision affirms a series of cases in which the ECJ has given weighty recognition of the basic rights to privacy and data protection in the European Union’s legal system and upholds the principle of ‘proportionality’ as the most important principle in judicial review of acts which interfere in the private life of an individual as set forth in the Rundfunk Case Judgment. [xxix] The court acknowledges that the basic privacy rights are separate and distinct from the basic rights to data protection and examines the measures established by DRD as a distinct interference to the two rights. However, this decision does not support complex and vague arguments about the difference between “private data” and “private space”. [xx]

The validity of the data used in case disclosure must be guaranteed so that the activities carried out do not damage the sense of social justice in the community. This is in line with the basic philosophy of sociological justice which explains that a normative entity is committed to delivering humans to a just, prosperous, and happy life. [xxv]

The first one is the use of artificial intelligence or AI in detecting crime schemes. One of the examples is the success of artificial intelligence technology or AI in detecting fraudulent crime schemes that take advantage of the weaknesses of digital financing or fintech. Cybercriminals try to apply for loans using other people’s data.
This is detected by the Cashwagon submission system. The system, which consists of AI, Machine Learning/Internet of Things, and Big Data, finds suspicious applications by syndicates to cheat. Thus, the use of AI has shown its success in uncovering criminal cases.

Next, the uses of the Internet of Things and Big Data give birth to an IT-based application or method that facilitates the technical implementation of police officers in uncovering criminals. For example, in the case taking place on 4 December 2015, NYPD Officers at New York Police Station 73 received a warning on their cell phone from the NYPD’s new “Shot-spotter” system: Eight shots were fired near 409 Saratoga Avenue in the Bedford-Stuyvesant area, Brooklyn. The NYPD officers were able to carry out their duties quickly with the help of the Situational Awareness System, known as the Domain Awareness System (DAS). This system, developed based on sensor and signal technology, makes it easier for NYPD Police officers to collect and share data in carrying out their duties. [xxxii]

CONCLUSION

Big Data can be used appropriately in disclosing crimes that occur, both conventional crimes and cybercrimes. Disclosure of crimes using Big Data is closely related to the development of computing and data-storage technology. Its implementation requires a flexible legal institution so that it is not susceptible to continuous technological developments which cause changes in the mode of crimes.

The important benefits of Big Data and Computing Technology in the framework of investigations to uncover criminal cases are tracing data, storing data, and accessing records in digital traces through physical copies of old photos that no longer exist or are destroyed. Criminals may post incidents, photos, or events on social media that can be traced to reveal the crimes they committed. Big Data Analytics saves time finding and verifying the criminals’ posts. In the use of Big Data to reveal crimes, the users must not neglect the protection of the owners of the personal data. In preparing a regulation for disclosing criminal cases using Big Data, it is recommended that the preparation should be linked to the regulation on “Personal Data Protection”.

REFERENCES


Utilization of Big Data in Criminal Case Disclosure by the Indonesian Police Investigators


