

Big Data Analytics in law enforcement and crime prediction

Deepthy Prasad
depr5821@student.su.se



Abstract

The admiration of Big Data Analytics methods such as Machine learning, Artificial intelligence, Natural Language processing, etc by law-enforcement agencies is increasing as the traditional methods are less supportive in the analysis of the heterogeneous data. This research study investigates whether the BDA can support crime prediction from historical crime data and how the BDA supports the decision-making in a real-time law-enforcement organization. The research paper discusses some of the state-of-the-art BDA methods and their application using crime data through a short literature review and proves that BDA can be used in crime prediction. The study proposes an IT architecture based on BDA for a Police department which is a common law-enforcement agency in most countries. The practical application of BDA on crime data is also showcased in this research to show the promising results which BDA can assure in the decision-making in a law-enforcement organization.

Introduction

Background

The evolution of technology made drastic changes in human life and the internet became an indispensable part of our day-to-day life. A large amount of data becomes accessible through the internet from anywhere anytime. As the amount and sources of information have increased, the law enforcement organizations such as Police, Federal agencies, and various private and public investigation agencies also started to make use of the benefits of technological advancements in order to improve the efficiency of their services. These sources include structured and unstructured data which requires large memory such as video footage, audio files, chat logs, geodata, weblogs, etc. This type of data is known as Big Data and more accurately Big Data refers to large and complex datasets [1]. The traditional tools and approaches became incapable to handle this kind of heterogeneous data which lead to the need for new tools and methods [2]. ‘Big data analytics (BDA)’ is an effective and fast-forwarding way to manage the multi-sourced, unstructured, and huge volume of data compared to the traditional methods. ‘Big data analytics (BDA) is a systematic approach for analyzing and identifying different patterns, relations, and trends within a large volume of data’ [3].

There are many research papers discussing the application of Big Data analytics in various domains such as transportation, healthcare, law enforcement, marketing, etc. The research paper examines the potential of big data analytics for security intelligence and also provides a future research direction in enhancing criminal investigations using big data analytics. Another research paper analyses how the BDA can be used to enhance business intelligence and also proposes a big data analytics service-oriented architecture (BASOA). Another paper is a comparative study of a few big data analytics methods and how can they be applied to various operation management areas such as forecasting, inventory management, revenue management, transportation management, etc. The future study recommendations of this paper are the basis of our research.

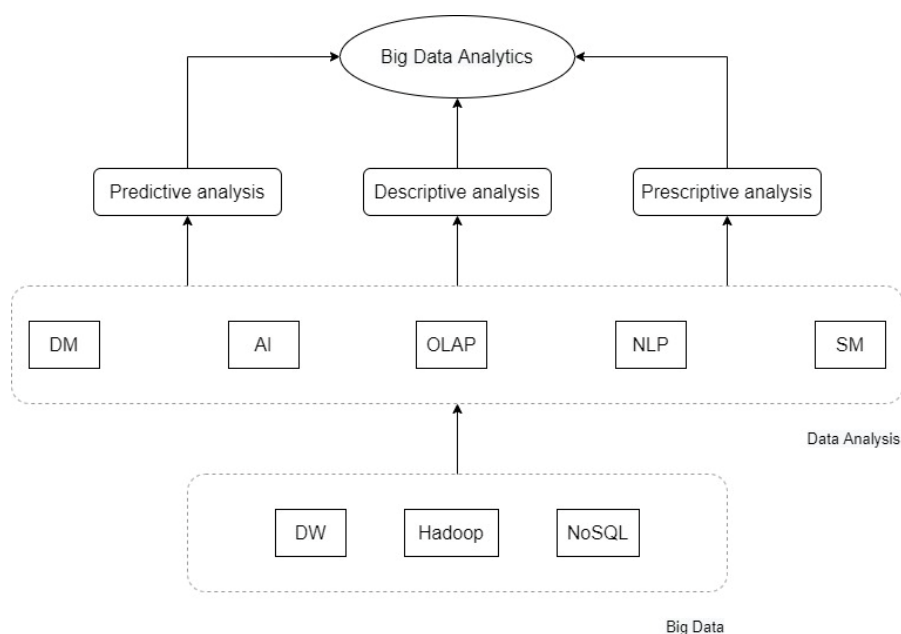


Figure 1. Big Data Analytics Ontology

This research study is based on the Big Data Analytics ontology in *Figure 1. Big Data Analytics Ontology* [4]. The ontology has three layers which include the components of BDA, various data analysis methods and different big data management systems. The main three components in BDA are descriptive, predictive, and prescriptive analytics. Descriptive analytics shows what has already happened, predictive analytics is used to predict future outcomes based on existing big data and prescriptive analysis shows what should happen with the best outcome [5], [6], [7]. All these components can be used together to produce the best results [8]. The middle layer of the ontology has some state-of-the-art data analytics methods for Big Data which include DM- Data Mining, AI – Artificial Intelligence, OLAP - Online analytical processing, NLP – Natural Language Processing and SM - statistical modelling. The bottom layer includes various data management systems such as DW - Data Warehouse, Hadoop and NoSQL databases which can handle a large amount of heterogeneous data.

Problem

Nowadays, law enforcement agencies often come across with large volume of crime data from multiple sources and they have realized that this can be utilized for investigation and crime prediction [9]. The traditional centralized data storage method also has been changed to distributed storage which supports large and complex datasets known as big data and its analysis. The better performance and cost benefits are some additional reasons that encourage the change from traditional to big data analytics [2]. A research paper by Tsan-Ming Choi, Stein W. Wallace and Yulan Wang discusses the application of Big Data Analytics in Operations Management where they have compared various existing big-data-related analysis techniques and also proposes future research in the areas such as Big data architectures and applying big data to real-world issues [10]. This research paper is based on this proposed future study where the application of big data analytics in law enforcement is discussed. This paper discusses how the real-world issues in law enforcement can be solved efficiently using big data analytics and propose an IT architecture for the implementation of big data analytics in the Police department.

The problem addressed in this essay is whether the big data analytics can be used in law enforcement to predict crimes and how BDA supports a real-time law-enforcement organization such as a Police department in their daily operations.

Research question

The research question to be answered in this essay are, whether the Big Data Analytics in law enforcement supports the crime prediction or not and how BDA supports a real-time law-enforcement organization such as a Police department in their daily operations.

Research method

The current paper is qualitative research to find out how big data analytics can be used in law enforcement organizations and predict crime. A research strategy should be selected in such a way that the study results are descriptive enough to draw inferences regarding the research question. After the analysis of various research strategies available for qualitative research, 'Record keeping' was found to be the most suitable strategy for the current study. "Record keeping method makes use of the already existing reliable documents and similar sources of information as the data source" [11]. Our research study also makes use of existing documents such as journals, articles, textbooks, etc on big data analytics tools and methods. Also, the corresponding data collection method used for the current study is 'document review' where existing data in form of various documents are collected and analysed to conduct the research. According to the website QuestionPro [12] "Document review is an efficient and effective way of gathering data as documents are manageable and are the practical resource to get qualified data from the past".

The sample for the research is collected from Google Scholar, ACM, ResearchGate, IEEE Xplore and Science Direct databases. The literature collected from these sources is provided in the reference section. A data evaluation is carried out in order to find the relevant literature and then analysed the data to draw inferences from pertinent literature. This research study has a theoretical part along with an architecture of the practical implementation of the conclusions in a real-time organization.

Theoretical base and concepts

This section includes a detailed description of the data analytics methods in the third layer of *Figure 1. Big Data Analytics Ontology* and the theoretical base for the application of these methods in crime prediction. A review of 5 research papers was also included in this study in order to find state-of-the-art analytics models and techniques used on Big Data for crime prediction.

Explanation of Big Data Analytics methods in the Ontology:

Data Mining (DM) is the process in which both structured and unstructured data are analyzed in order to find the patterns and relationships to support business intelligence and decision making. It uses various technologies including Artificial Intelligence (AI), Deep learning (DL), Statistical Modelling (SM) and Machine Learning (ML) for predictive, descriptive, and prescriptive analysis of data [13]. OLAP is a Data warehouse-based technology for the analysis of information. It supports the organization of data in a multi-dimensional model so that the business users can understand it easily and also provide capabilities to analyze the trend in data [14]. Artificial Intelligence refers to the ability of a machine to imitate human intelligence. So, AI makes use of various technologies such as ML, NLP, DL, etc [15]. Natural Language Processing (NLP) is a part of AI and supports the processing of human language in text and voice formats in order to implement easy communication between humans and machines [16]. SM is used for the statistical analysis of data where the results show the relationship between different variables in the data and support drawing inference and predictions from data. SM can provide a mathematical representation of data which can be used in the identification of crime series [17].

Literature Review:

No	Research Paper	The goal of the research	Data analytics methods	Findings	Future research
1	A machine learning-based forensic tool for image classification - A design science approach [18]	This research paper suggests how pre-trained models can be incorporated into the development life-cycle of an ML-based forensic tool for classifying the guns in the images	Keras pre-trained models: Convolutional Neural Network (CNN) model: <ul style="list-style-type: none">• Inception V3• Xception• VGG16• Residual Network (ResNet)	An ML-based digital forensic tool that utilises pre-trained models for a hand-gun classification even for the developers with less ML knowledge	Focus on evaluating real cases, like searching through the drugs and images related to drugs. Illicit drug image recognitions. Applying multi-criteria decision-making methods to compare the models.

2	Digital forensics supported by machine learning for the detection of online sexual predatory chats [19]	This paper implements machine learning methods to identify and classify harmful chat logs based on a digital forensic process model.	<p>Logistic Regression (LR)</p> <p>XGBoost</p> <p>Multilayer Perceptrons (MLP)</p> <p>Long Short-Term Memory (LSTM)</p>	<p>The study uses online data in XML format.</p> <p>Dimensionality reduction and decoding specific symbols to text form were part of the data preparation stage.</p> <p>Four types of classification models were used to train the data. Then performed testing to produce results and evaluation to measure the accuracy.</p> <p>LR, MLP and Bidirectional LSTM were found to be the best among used ML techniques.</p> <p>The study was able to classify the data and also found some words and phrases in harmful chat messages.</p>	This study can be made quantitative by using more data to train and validate the models proposed to generalise predatory behaviour.
3	Deep Learning-based Facial Detection and Recognition in Still Images for Digital Forensics [20]	The aim is to implement Face Detection and Recognition in Images (FDRI) into the digital forensic software Autopsy to speed up the digital forensic process	The FDRI algorithm implements a deep CNN to analyse the images in the dataset	<p>High precision of face detection can be achieved when using large enough images. Although each result is still highly dependent on specific features like illumination and face orientation.</p> <p>Photos used in identification documents are not ideal when it comes to applying facial recognition in everyday scenarios.</p>	As hardware evolution continues, the processing speed will increase and as databases grow images will have better training sets and could become more accurate.
4	Deep learning based static taint analysis approach for IoT software vulnerability location [21]	The aim is to develop a deep learning-based Internet of Things (IoT) vulnerability locator.	<p>Recurrent Neural Network (RNN)</p> <p>Long Short Term Memory (LSTM)</p> <p>Bidirectional LSTM (BLSTM)</p> <p>Convolutional Neural Network-based LSTM (CNN-based LSTM)</p>	<p>Used three taint selection principles to determine all the taints in the IoT software. Then used a taint weight calculation method to perform taint screening. 4 deep-learning-based models are used for vulnerability detection.</p> <p>The effectiveness of various models is evaluated using Code Gadget Database.</p> <p>The proposed CNN-based LSTM model showed higher accuracy than the other 3 models.</p>	Proposes future works where the proposed system can be applied at the industrial level to ensure the safety of IoT devices.
5	A Malware Detection Framework Based on Forensic and Unsupervised Machine Learning Methodologies [22]	The paper proposes a forensic malware detection framework based on unsupervised	K-means clustering algorithm.	Data containing malware Indicators of Compromise are collected and normalized the data. Cyber threat intelligence data malware name severity and state are used in this paper. Then the K-means clustering algorithm is	Improvement of the proposed framework with other ML techniques is suggested as future work.

machine learning
techniques.

used to find the pattern in the data
and find out the state with the
highest malware infections.

Table 1. Literature Review

Explanation of analytical Models and techniques used in 5 research papers that were reviewed
Table 1. Literature Review :

Regression analysis is a widely used machine learning technique to calculate the relationship between a dependent and several independent variables and its applications include prediction, forecasting, time-series data analysis, etc. Logistic Regression (LR) is an algorithm used for predicting binary or multi-class classification problems [23]. Gradient Boosting is another classification technique in ML which provides an ensemble of weak prediction models in order to reduce the prediction error. Extreme Gradient Boosting (XGboost) is an advanced form of gradient boosting to improve the generalization capability of the model [24]. K-Nearest Neighbour (KNN) is a prediction algorithm which can be used in both regression and classification problems [25]. So that it will be beneficial in finding the values such as malicious codes, website names, etc in crime data analysis. The decision tree is another basic algorithm in supervised learning for classification which is the basis for other popular ML algorithms such as Random Forest, Gradient boosting, etc [25]. If the features of the data are uncorrelated, the classification algorithm ‘Naive Bayes’ also can be used but the accuracy may vary according to the type of data [25]. The neural network is a popular method used in Artificial intelligence to find the patterns from the input data by processing it in a way that the human brain works [26]. Major divisions of the NN are Artificial Neural Networks (ANN), Convolution Neural Networks (CNN) and Recurrent Neural Networks (RNN). Neural networks can be used for both classification and pattern recognition [26]. The major application of NN in criminal investigation is the identification of fraudulent bank transactions, persons, websites, online predators through chat logs, etc. ANN helps the criminal investigations in the case of tabular, image and text input data whereas RNN is best suitable for time series data such as chat logs, audios, etc. CNN works well on image data to extract features which can be used in the face recognition of criminals. Multilayer Perceptrons (MLP) is a feed-forward artificial neural network (FFNN) model used for both classification and regression [27]. Long Short-Term Memory (LSTM) is an RNN used to make predictions based on time series data. LSTM is often used in Natural language processing which is a useful method for malware detection in forensic crime analysis [28].

Practical application

A Police department is a real-time law enforcement organization found in most countries. An IT architecture based on big data analytics for a Police organization is created to showcase the practical application of the theoretical part of this research study *Figure 2. IT architecture*. Then a research paper ‘Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data’ by Mingchen Feng, Jiangbin Zheng, Jinchang Ren, Amir Hussain, Xiuxiu Li, Yue Xi and Qiaoyuan Liu is presented as a proof of the real-time application of BDA in crime investigation [3].

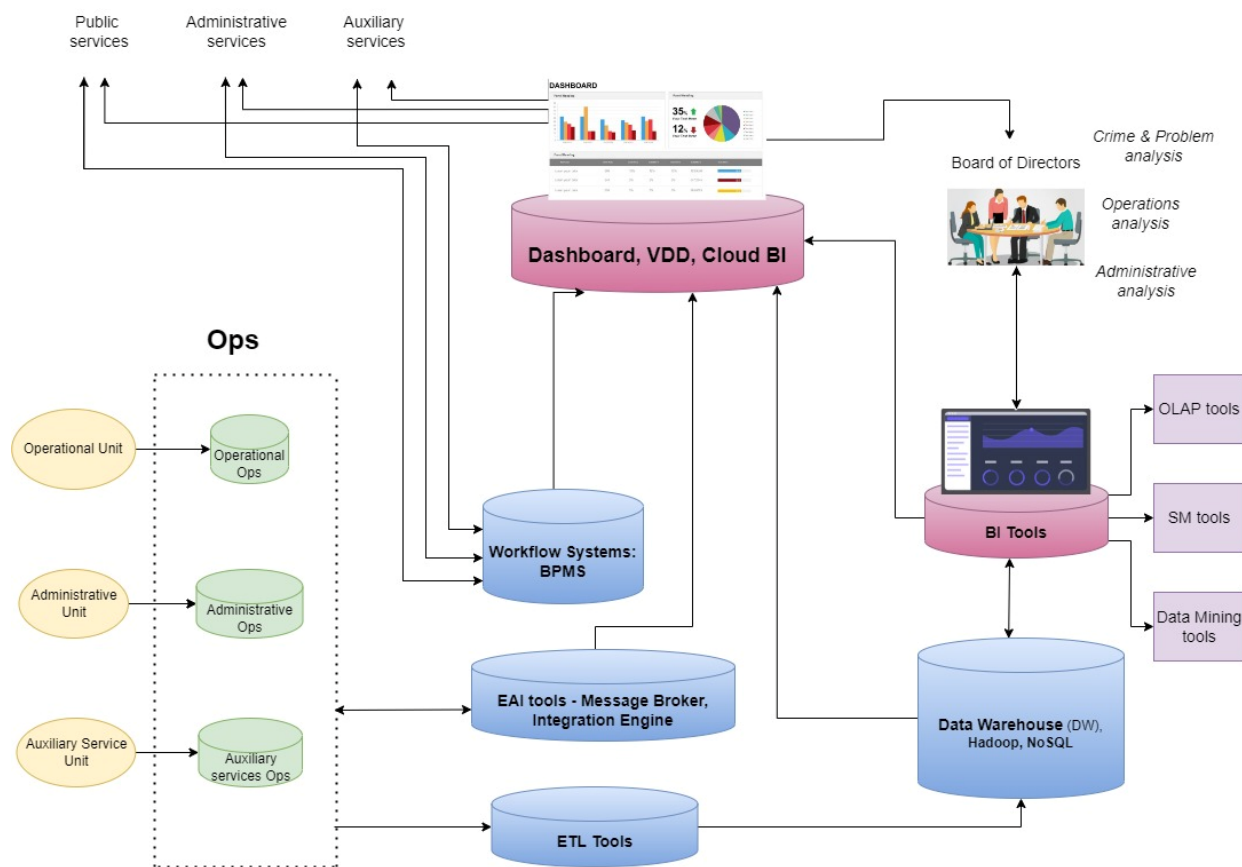


Figure 2. IT architecture

Organization described

The major operational systems that serve as the data input sources in the IT architecture of police organizations are Operations, Administration and Auxiliary service sub-systems these three sub-systems are interrelated and the basis of various operations and tasks in a police department. The operations subsystem performs tasks to assist the public directly such as traffic control, patrolling, criminal investigations, community services, etc. The administration department is responsible for long-term services required for the persons in the operations department such as budgets and finance, inspections,

implementation of business intelligence, information processing, etc. The auxiliary services take care of regular and ongoing tasks such as equipment and supply, telecommunications, maintenance, etc [29]. The data coming from all these subsystems are analyzed in order to assess the performance of each department because it determines the overall performance of the organization.

Explanation of IT architecture:

The operational systems (Ops) are the basis of the IT architecture of an organization as it provides all the information required for all IT systems. The major three subsystems ‘Operations, Administration and Auxiliary service’ of a Police organization are considered the operational systems in the IT architecture. ETL tools are used to extract, transform, and load the data from these subsystems to the databases such as DW, Hadoop, NoSQL, etc. EAI tools such as integration engine and message broker enable the communication and integration of the three subsystems. The workflow system ‘Business process management System (BPMS)’ optimizes and accelerates business processes to improve organizational efficiency. Business Intelligence tools are used to run the crime and problem analysis, operations analysis, and administrative analysis to find out the crime patterns, predict crimes and collect the evidence of the crime. The proposed IT architecture uses Big Data analytics tools such as OLAP, Data mining, and SM tools whereas the data mining tools use technologies such as AI, ML, and NLP.

Practical application

Considering the ethical constraints and time limitations of the current research, an existing study with the practical implementation of big data analytics (BDA) techniques in Police department tasks is reviewed and discussed in this section.

The research paper [3] uses publicly available datasets of 3 famous United States cities San-Francisco, Chicago, and Philadelphia to implement Big Data Analytics to identify various patterns, trends, and relations in a large volume of crime-related data. Selection of relevant features from the dataset, missing value imputation and discretization of time columns are carried out as part of the data preprocessing. The 13 features chosen were as follows [3]:

1. IncidentNum - Case number of each incident.
2. Dates - Date and timestamp of the crime incident.
3. Category - Type of the crime. This is the target/label that we need to predict in the classification stage.
4. Descript - A brief note describing any pertinent details of the crime.
5. DayOfWeek - Day of the week that crime occurred.
6. PdDistrict - Police Department District ID where the crime is assigned.
7. Resolution - How the crime incident was resolved (with the perpetrator being, say, arrest or booked).
8. Address - The approximate street address of the crime incident.
9. X - Longitude of the location of a crime.
10. Y - Latitude of the location of a crime.
11. Coordinate - Pairs of Longitude and Latitude.
12. Dome - whether crime id domestic or not.
13. Arrest - Arrested or not.

Initially, an exploratory analysis of data is used to visualize the basic time series and pie charts according to the type of crime and draw some basic inferences. The marker cluster algorithm [30] is used to cluster the crime incidents according to its latitude or longitude information and create an interactive map using this. This map showed the distributions of crime incidents in each city and found more crimes in downtown and fewer crimes on the border sides of the city. Then some of the big data analytics techniques such as the time series prophet model, traditional neural network, convolutional neural networks (CNN), and Recurrent neural network (RNN) - Long Short-Term Memory (LSTM) model are used to predict the future crime data for the three cities. The accuracy of each model is calculated using Root Mean Square Error (RMSE) [31] and Spearman correlation [32], the prophet and LSTM models show the maximum accuracy. A graphical representation of prediction using the existing data is included in the research paper which shows the monthly variation in the frequency of crime occurrences. Then a comparison of results from different layers in the neural networks is executed using RMSE and Spearman correlation. The study also provides the optimal parameters for both prophet and LSTM models to obtain maximum accuracy in the prediction.

Practical application discussed

The practical application of some of the state-of-the-art big data analytics techniques over crime data of 3 major US cities was able to showcase the graphical representation of crime forecasting. These results assure that this type of BDA prediction model can be used to find out the parts of the cities affected by more crime rates and also the peak times of crime rate in different areas of the city. The contributors of the research paper [3] suggest assigning the police force in different areas of the cities according to the results of the study to achieve the maximum efficiency of police department operations. The comparison results of predictions over different time periods data concluded that 3-year was the optimal period for the training sample. The LSTM and Prophet models were found to be more efficient in crime prediction compared to CNN and traditional neural networks such as ANN and perceptron, but they can vary according to the type of crime data. This practical application of BDA techniques is a shred of evidence that the BDA models perform different with each crime data and its accuracy needs to be measured at different levels to obtain the best crime prediction model and also the optimal parameters.

How to make the essay a research contribution

The current study can be considered as qualitative research of an introduction of BDA for law enforcement as it describes the basic ideas in BDA and proposes a BDA-based IT architecture for a law enforcement organization which is a Police department. This study can be further developed into a quantitative research study to measure the level of BDA usage in law enforcement organizations nowadays. Another application of the concepts introduced in this study is the BDA in other organizations such as healthcare, social service, etc. So, the mentioned quantitative research can be on any type of organization where they are currently using some of the BDA tools for business intelligence. According to Denscombe [33], The research strategy ‘survey’ will be a suitable option to do quantitative research as the survey can solve a problem using the information collected from a sample of individuals through the responses to various questions. Ordinal data is always suitable for quantitative research, so a questionnaire with closed-end questions can be used to collect data which can then be used in both descriptive and exploratory analysis of the data. The questionnaire can allow a large sample to participate in the data collection and thus more accurate results can be drawn from the data. The granularity and quality of collected data can be increased by using random sampling where the study population includes only relevant people. The questionnaire should be formatted in such a way that the inferences regarding the level of the business impact of BDA tools are obtained as results.

Conclusion

This research paper has discussed various state-of-the-art Big Data Analytics technologies and their application in law-enforcement organizations. Also proposed a BDA-based IT architecture for a Police organization. Also included a literature review in the theoretical concept section which provided some inferences on the various BDA methods used in real-time scenarios to collect evidence for and predict various criminal activities.

Research question revisited

The research question investigated in this research is whether the Big Data Analytics in law enforcement supports the crime prediction or not and how BDA supports a real-time law-enforcement organization such as a Police department in their daily operations. The literature review included in the theoretical part proves that the BDA supports crime prediction from a large amount of data and shows that the classification algorithms such as Neural Networks (NN), Decision tree, K-Nearest Neighbor, etc are the most commonly used big data analytics models in crime investigation as it is helpful to identify illegal

activities, face recognition of criminals and more often the proof of the crime that happened. The practical implementation part of this study indicates that the big data analytics methods allow the business analyst to extract relevant information from the huge amount of data and are also found to support crime forecasting along with the investigation in real-time organizations. These results can provide insights to Police departments to manage efficient resource allocation after tracking the criminal activities, and support optimized decision-making. The proposed IT architectures show that the BDA needs to be added as BI tools to support the organization in decision making.

Future research

The limitations faced in this research study were the lack of time and resources to complete thorough research on the topic. By having more time and man-hours, more papers can be found, analyzed, and compared on deeper levels. A future study can be conducted by reviewing more literature on BDA application in the law-enforcement area to find out more suitable methods for different criminal activities. Also, there is a scope of quantitative research in this area to measure the level of usage of BDA in current organizations. Because quantitative research is essential to prove the impact of BDA in decision-making and business intelligence. Another future research which can be suggested based on the current study is comparing different BDA techniques for different types of criminal data in order to find out the best prediction models and optimal parameters for the models.

Several words (excluding the words in the abstract and reference list): 4053

References

- [1] "Big Data," [Online]. Available: https://en.wikipedia.org/wiki/Big_data.
- [2] "big-data-vs-traditional-data," [Online]. Available: <https://treehousetechgroup.com/big-data-vs-traditional-data-whats-the-difference/>.
- [3] J. Z. J. R. A. H. X. L. Y. X. Q. L. MINGCHEN FENG, "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8768367>.
- [4] Z. Sun, L. Sun och K. Strang, "Big Data Analytics Services for Enhancing Business Intelligence," [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/08874417.2016.1220239>.
- [5] "Books_1," [Online]. Available: [https://books.google.se/books?hl=en&lr=&id=HYYaX9dsZsYC&oi=fnd&pg=PR13&dq=Mine lli+M,+Chambers+M,+Dhiraj+A.+Big+data,+big+analytics:+emer%3Fging+business+intelligence+and+analytic+trends+for+today%E2%80%99s+busi%3Fnesses.+Wiley+%26+Sons+\(Chinese+Edition+20](https://books.google.se/books?hl=en&lr=&id=HYYaX9dsZsYC&oi=fnd&pg=PR13&dq=Mine+lli+M,+Chambers+M,+Dhiraj+A.+Big+data,+big+analytics:+emer%3Fging+business+intelligence+and+analytic+trends+for+today%E2%80%99s+busi%3Fnesses.+Wiley+%26+Sons+(Chinese+Edition+20).
- [6] "Data, information and analytics as services," [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167923612001558>.
- [7] "Books_2," [Online]. Available: https://books.google.se/books?hl=en&lr=&id=4JN4CgAAQBAJ&oi=fnd&pg=PP1&dq=Database+systems:+design,+implementation,+and+management+&ots=VMusdqPa0a&sig=ZQmBPTyJ-sL0Mf2_iICdmYoT3zA&redir_esc=y#v=onepage&q=Database%20systems%3A%20design%2C%20implementation%2C.
- [8] "3-types-business-analytics," [Online]. Available: <https://online.bath.ac.uk/content/descriptive-predictive-and-prescriptive-three-types-business-analytics>.
- [9] "Introduction to Social Network Analysis (SNA) as an investigative tool," 2009. [Online]. Available: <https://link.springer.com/article/10.1007/s12117-008-9057-6>.
- [10] S. W. W. W. Tsan-Ming Choi, "Big Data Analytics in Operations Management," [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.12838>.
- [11] "Research startegy," [Online]. Available: <https://www.questionpro.com/blog/qualitative-research-methods/>.
- [12] "data-collection-methods," [Online]. Available: <https://www.questionpro.com/blog/quantitative-data-collection-methods/>.
- [13] "DM explained," [Online]. Available: <https://www.javatpoint.com/types-of-data-mining>.

- [14] "OLAP Explained," [Online]. Available: <https://core.ac.uk/download/pdf/82646817.pdf>.
- [15] "AI Explained," [Online]. Available: <https://chethankumargn.medium.com/artificial-intelligence-definition-types-examples-technologies-962ea75c7b9b>.
- [16] "NLP Explained," [Online]. Available: <https://www.ibm.com/cloud/learn/natural-language-processing>.
- [17] "SM Explained," [Online]. Available: <https://phys.org/news/2015-08-statistical-police-crime-series-apprehension.html#:~:text=0-Statistical%20model%20helps%20police%20identify%20crime%20series%3B%20speeds%20apprehension%20of,unveiled%20today%20by%20Michael%20D..>
- [18] "A machine learning-based forensic tool for image classification - A design science approach," 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666281721001827>.
- [19] "Digital forensics supported by machine learning for the detection of online sexual predatory chats," 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666281721000032>.
- [20] "Deep Learning-based Facial Detection and Recognition in Still Images for Digital Forensics," 2019. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/3339252.3340107>.
- [21] "A deep learning based static taint analysis approach for IoT software vulnerability location," 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S026322411931005X>.
- [22] "A Malware Detection Framework Based on Forensic and Unsupervised Machine Learning Methodologies," 2020. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/3384544.3384556>.
- [23] "Logistics Regression," [Online]. Available: <https://link.springer.com/book/10.1007/978-1-4419-1742-3?noAccess=true>.
- [24] "XGBoost," [Online]. Available: <https://towardsdatascience.com/a-brief-introduction-to-xgboost-3eae2e3e5d6>.
- [25] "ML Algorithms," [Online]. Available: <https://towardsdatascience.com/15-must-know-machine-learning-algorithms-44faf6bc758e>.
- [26] "neural-networks," [Online]. Available: <https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/>.
- [27] "Multi Layer Perceptrons," [Online]. Available: https://books.google.se/books?hl=sv&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&ots=F3WSdWaovi&sig=GhNskWIjQrjkCbc4Sb9gNJYLSLg&redir_esc=y#v=onepage&q&f=false.
- [28] "LSTM," [Online]. Available: <https://www.techopedia.com/definition/33215/long-short-term-memory-lstm>.

- [29] "Police organization," [Online]. Available: <https://quizlet.com/190431183/police-administration-chapter-4-police-organizational-tasks-flash-cards/>.
- [30] "marker cluster algorithm," [Online]. Available: <https://rsaiconnect.onlinelibrary.wiley.com/doi/abs/10.1111/pirs.12094>.
- [31] "RMSE," [Online]. Available: <https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e>.
- [32] "spearman correlation," [Online]. Available: <https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php#:~:text=Spearman's%20correlation%20measures%20the%20strength,is%20monotonic%2C%20but%20not%20linear..>
- [33] "Denscombe textbook," [Online]. Available: <https://www.researchgate.net/file.PostFileLoader.html?id=582a0dbf217e20276533f5a5&assetKey=AS:428404664213506@1479151039119>.

List of Figures and Tables

Figure 1. Big Data Analytics Ontology	2
Figure 2. IT architecture	8
Table 1. Literature Review	7

List of Abbreviations

BDA	: Big Data Analytics
DW	: Data Warehouse
BI	: Business Intelligence
ETL	: Extract Transform Load
OLAP	: Online analytical processing
BPMS	: Business Process Management System
VDD	: Visual Data Discovery

EAI	: Enterprise application integration
ML	: Machine Learning
AI	: Artificial Intelligence
DM	: Data Mining
DL	: Deep Learning
SM	: Statistical Modelling
NLP	: Natural Language Processing
NN	: Neural Networks

Department of Computer and Systems Sciences
 Stockholm University
 Forum 100
 SE-164 40 Kista
 Phone: 08 – 16 20 00
www.su.se

