

A Novel Approach for Deciphering Big Data Value Using Dark Data

Surbhi Bhatia* and Mohammed Alojail

Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University,
Al Hasa, 36362, Saudi Arabia

*Corresponding Author: Surbhi Bhatia. Email: sbhatia@kfu.edu.sa

Received: 10 September 2021; Accepted: 08 December 2021

Abstract: The last decade has seen a rapid increase in big data, which has led to a need for more tools that can help organizations in their data management and decision making. Business intelligence tools have removed many of the obstacles to data visibility, and numerous data mining technologies are playing an essential role in this visibility. However, the increase in big data has also led to an increase in ‘dark data’, data that does not have any predefined structure and is not generated intentionally. In this paper, we show how dark data can be mined for practical purposes and utilized to gain business insight. The most common type of dark data is a log file generated on a web server. Using the example of log files generated by e-commerce transactions, this paper shows how residual data and data trails can prove to be valuable when an actual dataset is inaccessible, and explains the usage of residual data for modeling purposes. The work uses a system identification approach, based on natural language processing for log file tokenization and feature extraction. The features are then embedded into the next step, which uses a deep neural network to identify customers for targeted advertising. The results achieve a significant accuracy and show how dark data has the potential to deliver value for business. Locating, organizing, and understanding dark data can unlock its relevance, usefulness, and potential monetization, but it is important to act when the benefits of use outweigh the costs of access and analysis.

Keywords: Dark data; big data; internet of things; information storage; enterprise resource planning; deep learning; NLP

1 Introduction

‘Big data’ typically involves the accumulation of structured, semi-structured, and unstructured data from various sources and in different sizes, while ‘big data analytics’ refers to the use of analytics to obtain insight from such data. The continuous expansion of information technology and the rapid increase in data capture related to the Internet of Things has led to an increase in big data. An enormous amount of data is generated by various sensors, wearable devices, and implementations of the Internet of Things. The amount of ‘excess’ data generated by these various sources, and the costs involved in its storage, will become an increasingly severe problem for all organizations in the coming years. The variety of data sources also increases the complexity of data management, and there are various risks when data is not understood and managed



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

appropriately. However, traditional database techniques have a limited application to big data because of its size. Obtaining meaning from big data by applying various analytical tools and techniques may prevent such data from accumulating in storage and avoid its conversion into ‘dark data’ [1].

Companies accumulate a vast amount of data through their day-to-day business operations and in dealing with customers. Much of this data is not used in their daily operations, yet most companies still store it. Companies usually retain such data for compliance purposes, and there is also a myth that the data may be useful at some point in the future and that its deletion may create problems. It has been reported that 80% of the data stored by data-driven organizations in 2019 was occupied by ‘garbage’ or ‘dark data’, and the proportion of unstructured data rose to 90% in 2020, as shown in Fig. 1. Dark data refers to the data that is generated as a result of an organization’s daily operations and the various processes that are applied to data input, leaving new content that is dormant and inactive in repositories, hidden in systems and servers, and underused or forgotten. Dark data is both hidden data that is inherited from system processes, and actual data whose potential value is not recognized.

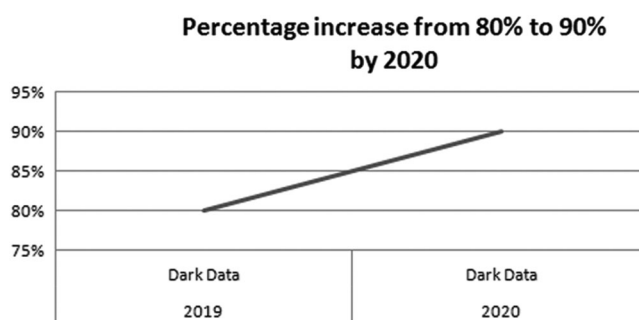


Figure 1: The rise of dark data

Dark data is not considered an information asset in an organization’s day to day operation of collecting, processing, and storing information. However, dark data could be utilized for other purposes such as analysis, business relations, and direct monetizing. The storing and securing of dark data typically involves more investment in order to obtain new ideas and helpful insight. The gains could be realized easily however by retrieving this kind of data and unlocking its hidden value for further use. If a company is paying for data storage but not using the data, the result is a wastage of finances and also of the opportunities related to that data [2]. Many companies have been unable to extract the value hidden in dark data because of a lack of proper tools. Additionally, companies that are resistant to change may find it difficult to keep pace with new challenges, so their business solutions cannot take full advantage of dark data [3]. While many business analysts think that dark data has colossal value, how to utilize it remains an unanswered question. Research has shown however that the strategy of following actual data-driven policies and implementing artificial intelligence (AI) can extract the value of dark data [4]. Dark data and AI now go hand in hand for a variety of companies [5]. The strategies for utilizing dark data should be used by every organization that deals with data.

For example, e-commerce businesses use targeted advertising by capturing phone numbers or email addresses which are then used as identifiers. Targeted offers can then be sent to the customer based on their purchasing behavior and adverts driven by machine learning. Whereas the front-end of the system is fully transparent to the software designer, the customer subjected to such an advert sees the process only as a “black box”. Dark data is accumulated from the variety of data that the system generates as the residue of ongoing processes. Usually, these datasets are treated as garbage and are auto-deleted from the system on a fresh boot. However, utilizing such datasets can convert the black box into a “grey box”,

which upon further analysis can expose the entire machine learning framework of a system from its output only. A simple schematic of the accumulation of dark data is shown in Fig. 2.

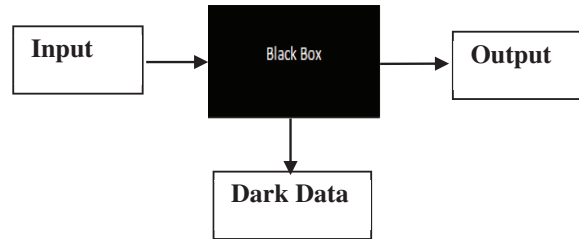


Figure 2: Schematic diagram showing the accumulation of dark data

Systems and servers usually generate log files, which are a common form of dark data. Log files are generated unintentionally during any process and contain raw and unstructured system configuration information. These log files can be read and understood by the use of natural language processing (NLP) methods. The use of classical NLP would be a difficult task because the amount of uncertainty is too large to handle. Therefore, in this paper we analyze log files using deep learning based NLP methods to develop a system for log mining and analysis and to replicate correctly the results of a black box machine learning model. When state of the art NLP techniques are used in conjunction with machine learning or deep learning methods, such methods are called “Modern NLP”. In this paper, dark data is used for model estimation and the generation of correct output. The correct output in this case is the customer who is to be targeted, represented by a label with customer name as header and contact details inside the label. Even if the user of the model is unaware of the machine learning algorithm, the results can still be replicated and an achievable accuracy achieved by analysis. The proposed algorithm has been implemented on the SAP R3 log file generated when using an Ensemble-based Ensemble (EBE) learning model for identifying customers for targeted advertising. The effectiveness of the proposed algorithm is tested using accuracy, precision, and percentage error calculations, and its effectiveness compared with the original algorithm.

The rest of this paper is organized as follows. Section 2 explains the importance of dark data, the benefits in leveraging it, and the challenge of managing it. Section 3 explains the methodology behind the proposed algorithm. Section 4 evaluates and discusses the test results, and Section 5 concludes.

2 Background

The world is interconnected, and data is a critical part of this interconnectedness. Companies rely on existing data to run their business operations efficiently. At the same time, new data is generated in almost every work instance. As explained in the introduction, the rapid increase in data has given rise to an increase in dark data. Cloud management companies have claimed that 52% of the data stored on their servers is dark data [6]. The phenomenon of dark data has become a challenge for data management, as well as an opportunity for businesses. Every company collects, processes, and stores data in the course of its day-to-day activities, but most companies fail to use this data for other purposes [7,8]. Many businesses are unaware of the value of dark data, and the possibility of monetizing such data [9]. Mining dark data means taking action to obtain useful information from such data and preventing a business from suffering a severe loss [10]. Some examples of dark data are given below.

- Log files generated by web servers.
- Archived data stored on the cloud.

- Video footage from surveillance systems.
- Data related to email correspondences.
- Data related to downloaded email attachments that have not been deleted [8].
- Customer-related information that is neither active nor used, such as call records [10].
- Information related to employees who no longer form a part of an organization.
- Information related to previous transactions, invoices, financial statements and accounts.
- Digital copies of notes, presentations, and old documents.
- Unstructured data collected by an organization such as raw survey data and unused internal data.
- Geo-location data.
- Data related to firewalls.
- Data related to systems architecture.
- Data generated from various advanced technologies such as machine learning.
- Data related to analytics programs.

Dark data could also be called ‘dusty data’. Some data repositories are full of data that is not structured, tagged, or even tapped, and is therefore “gathering dust”. The dark data in these repositories has not been processed or analyzed for business intelligence purposes, or to draw specific findings that could help in decision making [11]. The data has a similar value to big data but the difference is that organizations and analysts fail to realize this value [12–14]. When dark data is seen for the first time, it may not be considered very relevant. However, dark data incorporates a wealth of documents such as business proposals, account details, and email correspondence, all of which may be highly relevant at a later stage. For example, archived data may contain a business proposal intended to attract a client who might have rejected the proposal for some reason. After some years, the exact project details are required for consultation so that a new project can be presented to the client. Dark data can prove to be helpful for a strategic purpose. The retaining of data from ancient documents can be considered an asset. The safe storage of dark data can be seen as beneficial for customer relationship management, customer complaints, and the history of services. Additionally, there are specific compliance issues that require organizations to retain logs and records that may be considered as dark data. Such records are no longer necessary for day-to-day operations but need to be maintained in their original state by law. This kind of data should be filed securely and kept safe from unauthorized access and potential risks, but can be made accessible and used if the data is needed for some purpose [15]. The security and safety of dark data is essential as it contains business-critical information which may be utilized for the wrong purpose [16].

There are various obstacles to understanding the hidden value of dark data. Dark data comes in complex formats, making it difficult to recognize and categorize. Analysis is therefore difficult and the process is expensive [17,18]. However, many companies have excellent content management and storage strategies and add basic functionality to underutilized data effectively. If an organization fails to utilize data effectively, then applying analytics will result in poor intelligence. The analytic process is shown in Fig. 3.

Business systems store data that is considered to be unproductive and without value, but there are various productive ways of using it if companies implement strategies to connect, organize, and analyze this dark data [19]. The appropriate analysis can show hidden patterns missed during conventional data analytics [10–21]. Munot et al. [22] explained the importance of dark data and discussed applications for its analysis. Mokhtar et al. [23] discussed the impact of IR 4.0 in education and research, and showed how a university could adapt to IR 4.0 and function in the big data environment by exploring dark data examples. Researchers have also discussed types of dark data by exploring hidden cybersecurity risks, and showed how companies can be proactive by managing the data in the right way to reduce those risks [24]. Researchers have also discussed the dark data in IoT devices and the use of AI in business

analytics, and showed how businesses could benefit by unveiling trends which could increase productivity and efficiency [25,26].

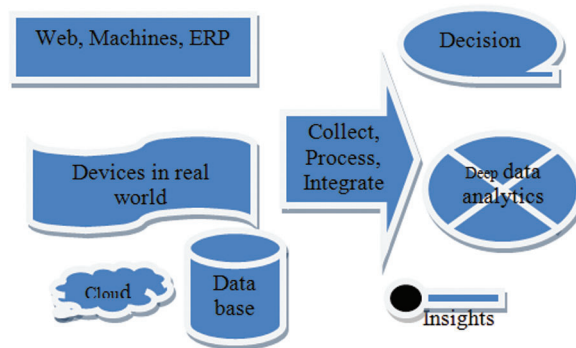


Figure 3: Dark data analytics

Dark data has specific properties which companies could utilize to release its potential as a business asset if they set up business models for handling it [9]. To have a competitive edge in the market, an organization should have the ability to recognize the relevance and usefulness of the data it holds, and adapt data to add value to the company or utilize it for the benefit of the outside world [18]. Knowledge comes from information, and there is a wealth of information in dark data. Information trails from credit card validations, delivery notes, and invoice details for instance, can all be used to identify product trends and selling opportunities, and to analyze ad hoc buying or spending patterns. In an automated payment process, information related to customers can be verified and checked from invoices at the first stage of the process so that data can be prevented from being converted into dark data. If the data is not analyzed, it will have cost implications as regards storage [19]. Data-driven companies need to train employees adequately so that meaningful insights can be extracted. The challenge of dealing with dark data is shown in Fig. 4.

Challenge dealing with Dark Data

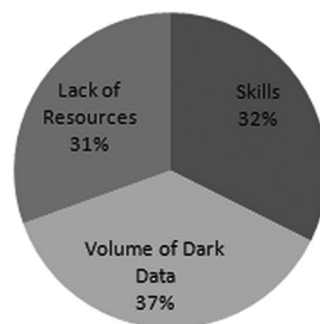


Figure 4: The challenge of dealing with dark data

3 Methodology

Log files are a type of data generated by web servers and system software, including ERP (Enterprise Resource Planning) software. Log files can be read and understood by the use of NLP methods. The files can be analyzed to identify various trends and utilized to avoid their conversion into dark data [15]. In this paper, we use dark data to model machine learning systems using ERP data and a system identification approach.

The traditional NLP approach is to develop a model that can handle four kinds of ambiguities: lexical, syntactic, semantic, and anaphoric. Because we are dealing with log files, we do not have such ambiguities and the design becomes straightforward. Stemming and lemmatization methods shall be used to convert the complex server/system logs into structured data [27,28]. Collecting log files from a server/system can either be done by Linux-based systems like Syslog and Logcat or by manually logging system activities. After the log file is generated, the data is preprocessed to take care of various errors and useless junk in the log, which reduces the size and workload of our NLP system. Using classical NLP would be difficult because of the large amount of uncertainty. Therefore, deep learning based NLP methods are used to develop an NLP system for log mining and analysis.

The proposed algorithm consists of two stages. In the first stage, NLP principles are applied to find the structure of the dataset. In the second stage, a primary deep neural network is used to find the desired output. Fig. 5 shows the proposed model.

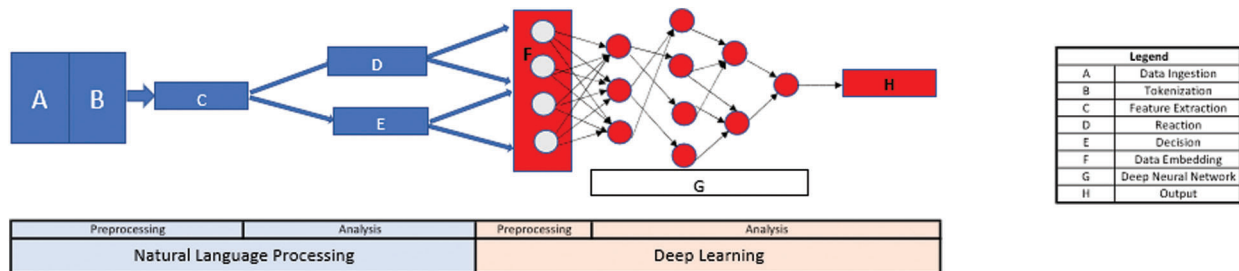


Figure 5: NLP-based deep learning model

Initially, the dataset is ingested into the NLP model [29] and the data is tokenized. These tokens prove to be helpful in the next step where the feature extraction is performed. In our study, we are aware that the model proposes action towards the targeted audience based on ERP data. Therefore, we expect two types of responses from the targeted audience. They will either react (labelled as “reaction”) or show interest/disinterest (labelled as “decision”). This is detected using a simple NLP-based classification model following feature extraction and completes the NLP stage of the algorithm. The features/classes thus identified are fed as an embedded dataset into a deep learning model. This is a basic deep neural network which has three layers, each with three nodes. A basic $\tanh(x)$ activation function has level_max as the “email” of the customer, level_min as the “phone number” of the contact, and level_mid as “unclassified”. Experimenting with other activation functions did not affect accuracy at the implementation stage. Later, to assess the model’s accuracy, the classified contacts from the NLP-based deep learning model, which analyzes log files to obtain its results, is compared with the results found by the EBE model that is run on the actual dataset. The process is summarized in Algorithm 1.

Algorithm 1: NLP-based deep learning methodology

1. Start
2. Define data: D
3. Ingest data into local DB
4. Tokenize, for (D = 1, D++, D <= End)
 - {
 - Get sub-word vocabulary dictionary (E): Names, Email ID;

(Continued)

Algorithm 1 (continued)

```

Get frequency of occurrence of E in D;
Get tokens, store tokens in F
}
5. Get frequency of frequency in F; Generate features G
6. Embed F into a deep learning model
7. Define hidden layers of deep neural network, target features to be defined as G
8. Activation function chosen as sigmoid, to either select or reject an output based on G
9. Output is a label which contains other information of contact

```

4 Results and Discussion

The work-based dataset is a log file from an e-commerce server that runs the EBE model for targeted audience campaigns. The dataset is not dynamic; however, the results of the blending process are different because different ensemble methods are employed in each process. The exact process is repeated by increasing the dataset. Brief information about each dataset is reported in [Tab. 1](#).

Table 1: Details of log files with types, attributes, and other details used as a dataset

Detail	No. of records in ERP dataset	No. of attributes	Type	Creator
Log file-1	150	4	Multivariate	SAP R3 Server, confidential
Log file-2	175	4	Multivariate	SAP R3 Server, confidential
Log file-3	300	4	Multivariate	SAP R3 Server, confidential
Log file-4	100	4	Multivariate	SAP R3 Server, confidential
Log file-5	125	4	Multivariate	SAP R3 Server, confidential

4.1 Lead Generation from EBE Model

Analysis was performed on the base ERP data shown in [Tab. 1](#). Two dummy datasets (named ‘iaculis’ and ‘blandit’) were included to ensure validity, showing transactions so that they get drawn into the algorithm. In the analysis based on the Ensemble-based Ensemble (EBE) model, the leads get caught by the algorithm and naturally also appear in the JSON log file (dark data) thrown out by the e-commerce server. For privacy, the base JSON (taken as the main file) cannot be shared. Because the log file was in JSON format, the NLP steps were rendered useless.

4.2 Lead Generation and Analysis Based on Deep Learning Back Propagation

Analysis was also performed on the log files. In this case, the dataset was embedded into a vanilla deep learning based neural network, using the base JSON data from the e-commerce site. The neural network has four layers, each with four input points. The output is only the email id and contact number for targeted advertising. The idea of using dark data in this process is validated if this analysis also catches iaculis and blandit for targeting. The process was repeated for up to 100 iterations to minimize errors, and the final result was output in the form of a CSV file. Both iaculis and blandit were observed in the CSV file.

The process was repeated again for up to 100 iterations to minimize errors, and the result again output in the form of a CSV file. Again, both iaculis and blandit were found in the results. Therefore, the analysis of the log files and the analysis of the actual data threw out the same results.

An accuracy level of 100% was achieved on the dummy data points. Overall, 29 data points were missed from the original prediction model. The model's accuracy was calculated based on the prediction of leads from the actual EBE model. The basic formulae are shown below. We label leads from the EBE model as A and leads from the dark data mining model as B.

$$Accuracy = \frac{(A) \cap (B)}{(B)} \times 100 \quad (1)$$

$$Error = (100 - Accuracy)/100 \quad (2)$$

$$Recall = \frac{True\ Positive\ [(A) \cap (B)]}{True\ Positive\ [(A) \cap (B)] + False\ Negative\ [(A) - (B)]} \quad (3)$$

$$Precision = \frac{True\ Positive\ [(A) \cap (B)]}{True\ Positive\ [(A) \cap (B)] + True\ Negative\ [(A) - (B)]} \quad (4)$$

The results of the metrics are shown in [Tab. 2](#). The metrics show that the accuracy rate of the synthetic data was 100% whereas the accuracy rate of the actual dataset was 79%.

Table 2: Accuracy, error, and recall scores calculated on different log file datasets

Detail	No. of records in ERP dataset	No. of customers targeted by the EBE model	No. of customers targeted by proposed model using server log files	No. of common customers in A & B	Accuracy (C/B)	Error	Recall	Precision
Dataset 1 (Log file)	150	25	36	29	80.56	19.44	0.44	0.68
Dataset 2 (Log file)	175	26	34	28	82.35	17.65	0.45	0.74
Dataset 3 (Log file)	300	34	31	24	77.42	22.58	0.43	0.60
Dataset 4 (Log file)	100	19	26	21	80.77	19.23	0.44	0.68
Dataset 5 (Log file)	125	21	27	20	74.07	25.93	0.42	0.65
Average accuracy				79.03				

[Fig. 6](#) shows the accuracy of the five log file datasets, which varies between 74% and 82%.

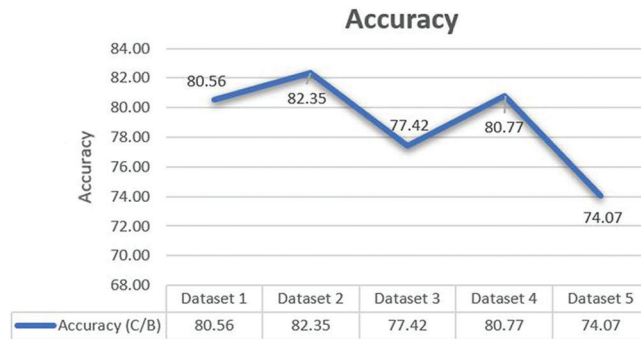


Figure 6: Accuracy rate calculated on five log file datasets

The error and recall scores are shown in Figs. 7 and 8, respectively.



Figure 7: Error rate calculated on five log file datasets

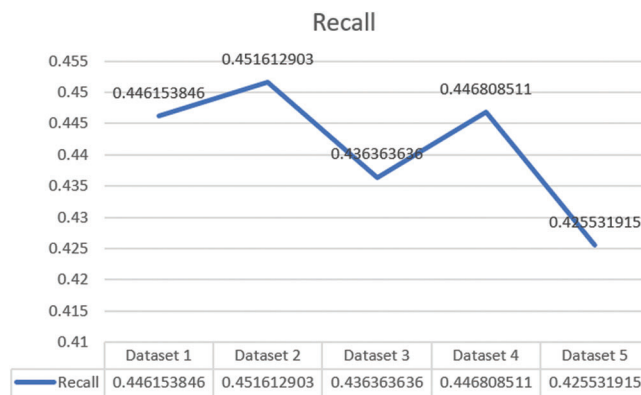


Figure 8: Recall scores calculated on five log file datasets

5 Conclusions

The last decade has seen a rapid increase in big data, which continues to increase on a daily basis. This increase has also led to an increase in dark data. When data is stored but not utilized, valuable data may become dark data and fall into the category of irrelevance or be treated as garbage. Data analysts argue that around 50%–60% of data is deprived of its value if analysis is not done immediately to avoid becoming dark data, which can cause a significant loss to an organization’s business model. In the case of

an e-commerce business for instance, if a group of customers are based in a particular location known to the company and the location can add some value to the business, then the data should be utilized immediately. Otherwise, the customer might change their location and the data becomes irrelevant. From a business perspective, dark data can help to provide insight, improve operations and critical decision making, highlight revenue opportunities, upgrade efficiency and quality issues, and increase productivity. In this paper, log files from an e-commerce server were utilized to run an EBE model for targeted audience campaigns. The results were analyzed and compared with the original model based on ERP data, and the results were found to be similar. This shows that even with a black box machine learning model, we can achieve similar results to the initial model with the help of dark data analysis. For business confidentiality reasons, this paper has not explored other examples of dark data such as financial statements and email correspondence using real-time data from large organizations. Future work needs to validate different deep learning algorithms on other examples of dark data, and develop a greater context for unveiling trends and relationships.

Funding Statement: This work was supported by the Deanship of Scientific Research, King Faisal University, Saudi Arabia grant number 216142.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] B. Schembera and J. M. Durán, "Dark data as the new challenge for big data science and the introduction of the scientific data officer," *Philosophy & Technology*, vol. 33, no. 1, pp. 93–115, 2020.
- [2] S. Few and P. Edge, "Big data, big ruse," *Visual Business Intelligence Newsletter*, pp. 1–8, 2012.
- [3] G. Gimpel, "Bringing dark data into the light: Illuminating existing IoT data lost within your organization," *Business Horizons*, vol. 63, no. 4, pp. 519–530, 2020.
- [4] J. P. Shim, A. M. French, C. Guo and J. Jablonski, "Big data and analytics: Issues, solutions, and ROI," *Communications of the Association for Information Systems*, vol. 37, no. 1, pp. 39–50, 2015.
- [5] D. Trajanov, V. Zdraveski, R. Stojanov and L. Kocarev, "Dark data in the internet of things (IoT): Challenges and opportunities," in *7th Small Systems Simulation Sym.*, Niš, Serbia, pp. 1–8, 12th-14th February 2018.
- [6] P. B. Heidorn, "Shedding light on the dark data in the long tail of science," *Library Trends*, vol. 57, no. 2, pp. 280–299, 2008.
- [7] C. Zhang, J. Shin, C. Ré, M. Cafarella, F. Niu *et al.*, "Extracting databases from dark data with deep-dive," in *Proc. of the 2016 Int. Conf. on Management of Data*, Association for Computing Machinery, New York, NY, United States, pp. 847–859, 2016.
- [8] K. Cukier and V. Mayer-Schönberger, "The rise of big data: How it's changing the way we think about the world," In: P. Mircea (ed.), *The Best Writing on Mathematics 2014*. Princeton: Princeton University Press, pp. 20–32, 2014.
- [9] D. J. Hand, *Dark data: Why What You Don't Know Matters*. Princeton: Princeton University Press, 2020.
- [10] C. Kimble and G. Milolidakis, "Big data and business intelligence: Debunking the myths," *Global Business and Organizational Excellence*, vol. 35, no. 1, pp. 23–34, 2015.
- [11] G. Eason, B. Noble and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 247, no. 935, pp. 529–551, 1995.
- [12] E. Ippoliti, "Dark data. Some methodological issues in finance," in *Methods and Finance*. Cham: Springer, pp. 179–194, 2017.
- [13] C. Patil and V. Siegel, "Shining a light on dark data," *Disease models & mechanisms*, vol. 2, no. 11, pp. 521–525, 2009.
- [14] L. Harrington, "New data of the digital age: Big, dark, and deep," *AACN Advanced Critical Care*, vol. 28, no. 3, pp. 239–242, 2017.

- [15] O. Soprano, "An Explorative Study on the Perceived Challenges and Remediating Strategies for Big Data among Data Practitioners," Master's Dissertation, Linnaeus University, Sweden, 2020.
- [16] B. Schembera and J. M. Durán, "Dark data as the new challenge for big data science and the introduction of the scientific data officer," *Philosophy & Technology*, vol. 33, pp. 1–23, 2020.
- [17] E. Letouzé, *Big Data for Development*. UN Global Pulse, Report, UN Global Pulse, Big Data for Development: Challenges and Opportunities, Unidas, Nueva York, mayo, 2020.
- [18] K. Migdał-Najman and K. Najman, "Big data: Dark data," *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, vol. 469, pp. 131–139, 2017.
- [19] Y. Liu, Y. Wang, K. Zhou, Y. Yang, Y. Liu *et al.*, "A framework for image dark data assessment," *In Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, Springer, Cham, pp. 3–18, 2019.
- [20] K. Munot, N. Mehta, S. Mishra and B. Khanna, "Importance of dark data and its applications," in *2019 IEEE Int. Conf. on System, Computation, Automation and Networking (ICSCAN)*, IEEE, Pondicherry, India, pp. 1–6, 2019.
- [21] M. Alojail and S. Bhatia, "A novel technique for behavioral analytics using ensemble learning algorithms in e-commerce," *IEEE Access*, vol. 8, pp. 150072–150080, 2020.
- [22] K. Munot, N. Mehta, S. Mishra and B. Khanna, "Importance of dark data and its applications," in *IEEE Int. Conf. on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, IEEE, pp. 1–6, 2019.
- [23] S. Mokhtar, A. Q. J. Alshboul and G. O. A. Shahin, "Towards data-driven education with learning analytics for educator 4.0," in *Journal of Physics: Conf. Series, Int. Conf. Computer Science and Engineering (IC2SE)*, Padang, Indonesia, pp. 2–8, 2019.
- [24] W. Dimitrov, S. Syarova and L. Petkova, "Types of dark data and hidden cybersecurity risks," in *Conceptual and Simulation Modeling of Ecosystems for the Internet of Things (CoMein)*, pp. 1–11, 2018.
- [25] G. Gimpel and A. Alter, "Benefit from the internet of things right now by accessing dark data," *IT Professional*, vol. 23, no. 2, pp. 45–49, 2021.
- [26] N. P. Rana, S. Chatterjee, Y. K. Dwivedi and S. Akter, "Understanding dark side of artificial intelligence integrated business analytics: Assessing firm's operational inefficiency and competitiveness," *European Journal of Information Systems*, vol. 17, no. 2, pp. 1–24, 2021.
- [27] S. Bhatia, M. Sharma and K. K. Bhatia, "Strategies for mining opinions: A survey," in *2015 2nd Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, IEEE, Delhi, India, pp. 262–266, 2015.
- [28] S. Basheer, K. K. Nagwanshi, S. Bhatia, S. Dubey and G. R. Sinha, "FESD: An approach for biometric human footprint matching using fuzzy ensemble learning," *IEEE Access*, vol. 9, pp. 26641–26663, 2021.
- [29] S. Priyadarshy, "Big data, smart data, dark data and open data: eGovernment of the future," in *2015 Second Int. Conf. on eDemocracy & eGovernment (ICEDEG)*, IEEE, Quito, Ecuador, pp. 16, 2015.