International Journal of
Environmental Research and Public Health

MDPI

Article

NLP-Based Digital Forensic Analysis for Online Social Network Based on System Security

Zeinab Shahbazi and Yung-Cheol Byun *

Department of Computer Engineering, Major of Electronic Engineering, Institute of Information Science & Technology, Jeju National University, Jeju 63243, Korea; zeinab.sh@jejunu.ac.kr
* Correspondence: ycb@jejunu.ac.kr

Abstract: Social media evidence is the new topic in digital forensics. If social media information is correctly explored, there will be significant support for investigating various offenses. Exploring social media information to give the government potential proof of a crime is not an easy task. Digital forensic investigation is based on natural language processing (NLP) techniques and the blockchain framework proposed in this process. The main reason for using NLP in this process is for data collection analysis, representations of every phase, vectorization phase, feature selection, and classifier evaluation. Applying a blockchain technique in this system secures the data information to avoid hacking and any network attack. The system's potential is demonstrated by using a real-world dataset.

Keywords: digital forensics; natural language processing; blockchain; machine learning

Citation: Shahbazi, Z.; Byun, Y.-C. NLP-Based Digital Forensic Analysis for Online Social Network Based on System Security. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7027. https://doi.org/10.3390/ijerph19127027

Academic Editors: Wajahat Ali Khan, Maqbool Hussain and Muhammad Afzal

Received: 7 April 2022
Accepted: 6 June 2022
Published: 8 June 2022

1. Introduction

Social media is generally used to communicate on the internet through various channels, to collaborate with different users, and to share information. The shared content supports researchers in investigating the potential of the criminal process. Social media does not have any limitations for content sharing related to victims, suspects, and witnesses [1,2]. The websites and applications are used to facilitate the sharing of content between connected networks. One of the social structures is the online social network (OSN), which which includes platforms such as Twitter or Facebook [3–6]. Forensic data extraction from social media platforms has become a considerable research problem [7–9]. Conventional digital forensics collects most of the information, which is a huge art of proof. Nevertheless, the extraction process is not practical on the OSN regarding the nature of the highly distributed network, shared content, and data size. Data collection from the individual subjects without any acceptable reason is almost unmanageable, and because of privacy laws, limited access is permitted [10–12]. Forensic data collection connects to the system operator for the formatting issue and data authenticity. The available digital forensics (DF) methods entail many challenges in cyber-physical systems. This includes the difficulties of data access, data originating from various locations, the traceability and transparency of evidence, and huge-volume data analysis. During the past few years, a large number of researchers have focused on forensic analysis based on cloud computing [13–15], evidence modeling [16–18], and assisting the community of law enforcement. Blockchain technology is a distributed ledger system that collects and saves the proper records in the decentralized format of a peer-to-peer network. The stored data are based on a timestamp block, and directly link with the chain based on proof of trust [19–21]. The advantages of applying blockchain in the DF system are to provide the digital evidence the accessibility of self-verification for ensuring the hash function and evidence chain verification. This process guarantees the system transparency, security, and immutability in case of examination. In this paper, we propose a digital forensic platform using the integrated method of NLP and

blockchain, feature selection using machine learning techniques, network vectorization, and system security analysis. Moreover, the presented system focuses on the relationship between content communication and individuals. The system uses the supervised NLP for topic extraction and applies the feature selection for topic ranking to find the highly weighted topics. Regarding the ranked topics, the classifiers can train with the famous algorithms and generators, and the output will be effective classifiers that can modify various metrics for further investigation. The main contributions of this paper are summarized as below:

- This research applies natural language processing techniques for the detailed data analysis approach;
- One of the important aspects of this research is the multi data source input, which makes this process more competitive with other research results;
- The main focus of this research is a system security method which stores the OSN information in blockchain framework.

The remainder of this process is summarized in Section 2, which reviews the recent related literature and the current state of the art. Section 3 presents the details of the proposed digital forensics approach and a performance evaluation. Section 4 presents details for the results, the implementation of the proposed digital forensics analysis, and finally, the conclusion.

2. Related Work

In this section, the state of the art in DF is presented in detail. The main focus is on two parts. One is digital forensics challenges in blockchain, and the other is the forensic attainment of social media content.

2.1. Digital Forensics Challenges in Blockchain

In DF, hash functions are applied to maintain digital integrity and generate the digital digest to avoid changes of digital assets [22,23]. Nevertheless, in the applications related to DF, the main focus is on disk drive integrity and the validation of data. The biggest concern is the hash validation and verification for special files such as images. The DF approach depends on the investigators' experience [24,25]. Some of the challenges related to the existing DF are presented as trustworthiness, integrity, provenance improvement, scalability, and availability [26,27]. In terms of trustworthiness, the system is supposed to check the trust if insider threats to the blockchain environment improve the trust of evidence [28]. Regarding integrity, the system checks the events examinations and items in the digital investigation. A traditional investigation provides forensic activities and supports the data, tools, etc. The improvement of provenance fetching at the top of hash functionality gives the information of hash validation to examine the system's behavior with creating the hash tree. The scalability in the hash tree is able to support nodes of the system, and it is capable of the hash digest in the deep level [29–31]. Every blockchain node contains the whole hash information, guaranteeing accuratcy. This aspect saves the digital data for investigating forensic events.

2.2. Forensic Attainment of Social Media Content

The DF attainment involves steps of proofing the criminal cases regarding location, security, data, etc. [32]. The data provided from social media is more understandable and easy to access for users [33–35]. To use this type of data, it is required to follow the legal and formal process [36]. This process is performed by a highly skilled person with sufficient knowledge of technical and legal matters [37–39]. The artifacts of DF identify the critical sources for social media evidence [40,41]. Thus, lots of research materials focus on the attainment of forensic evidence; the extraction of forensic information from social media concentrates on the identification of specific devices and detects the traces found by the device from the web browsers or media applications [42–45]. To collect the forensic data, the requirements are defined as relevant data collection from multiple websites, metadata

collection from social media information, and certifying the data integration in the forensic collection [46,47]. DF footage is mostly used for the comparative analysis of images and objects to find the relative subjects to provide the opinion findings [48–51]. Table 1 shows the existing state of the art in forensic data analysis. The main focus is on the research approach, limitations, and advantages analyzed in DF. The selected works show various DF analyses based on machine learning, video clustering, chat data encryption, instant message analysis, etc.

Table 1. Comparison of the recent state-of-the-art DF methods.

Author	Proposed Approach	Advantages	Limitations
Choi et al. (2019) [52]	Digital forensic analysis for the Kakao Talk encrypted data.	Data recovery without requiring a password from a user.	Difficult to protect user-sensitive information.
Zhang et al. (2018) [53]	Digital forensic analysis for smart phone instant messaging.	Investigating the history of user messages in four Android mobile applications.	Limitation in communication mode for one-to-one contacts.
Du et al. (2020) [54]	Future of artificial intelligence in investigation of digital forensics.	Survey of automated evidence-processing methods based on AI techniques.	Image data with low quality is difficult to train and process further.
Xiao et al. (2019) [55]	Analysis of video-based evidence investigation of digital forensics.	Identification of forensics and link-establishment to investigate the objects.	Difficult to identify human face recognition, motion detection, etc.
Jadir et al. (2018) [56]	Digital forensic enhancement for document clustering,	Enhancing the document clustering performance for partitioning the criminal reports and text dataset.	Challenges of processing if the data numbers increase.

3. Proposed NLP-Based Digital Forensic Analysis for Online Social Network

This section briefly presents the integration of NLP techniques with blockchain. Figure 1 shows the overview of the proposed digital forensic analysis in terms of the NLP and blockchain approach. The main goal of this system is to improve the security of the DF analysis regarding the information shared on social media.

NLP techniques are applied to analyze the collected dataset from every aspect to provide meaningful information to the proposed system. This process has five main layers: a processing layer, an interface layer, an analysis layer, a data layer, and a knowledge layer. The responsibility of the processing layer is to identify and acquire the system input. Inputs are from social networks and incident notifications for which identification requires the incident specification and incident boundary specification. The identification, extraction, data collection, the parser, and preservation are required before moving forward to acquisition. Completing this process, the forensic data is the input of the data layer. In this layer, the hybrid data mapping is used for global and local ontology and storing the data. The next layer is the analysis layer for which the interface query is an NLP semantic interface, and the analysis operators are correlation events and document analysis, location analysis, the relatedness of findings, frequency analysis, and relationship analysis. The analysis

report moves to the interface layer, which contains the timelines, tweet cloud, temporal graph, interaction graph, frequency chart, and location chart. Next is the knowledge layer, where the relationship between the extracted dataset and its linking is processed. After completing the NLP steps and data processing, the analyzed dataset is ready to save into the blockchain framework. The main reason for using the blockchain framework is to secure the collected dataset with limited access to avoid hacking or attack. The blockchain framework contains the data collection, investigation, and verification processes, which provides the verified data to a court for defense and prosecution.

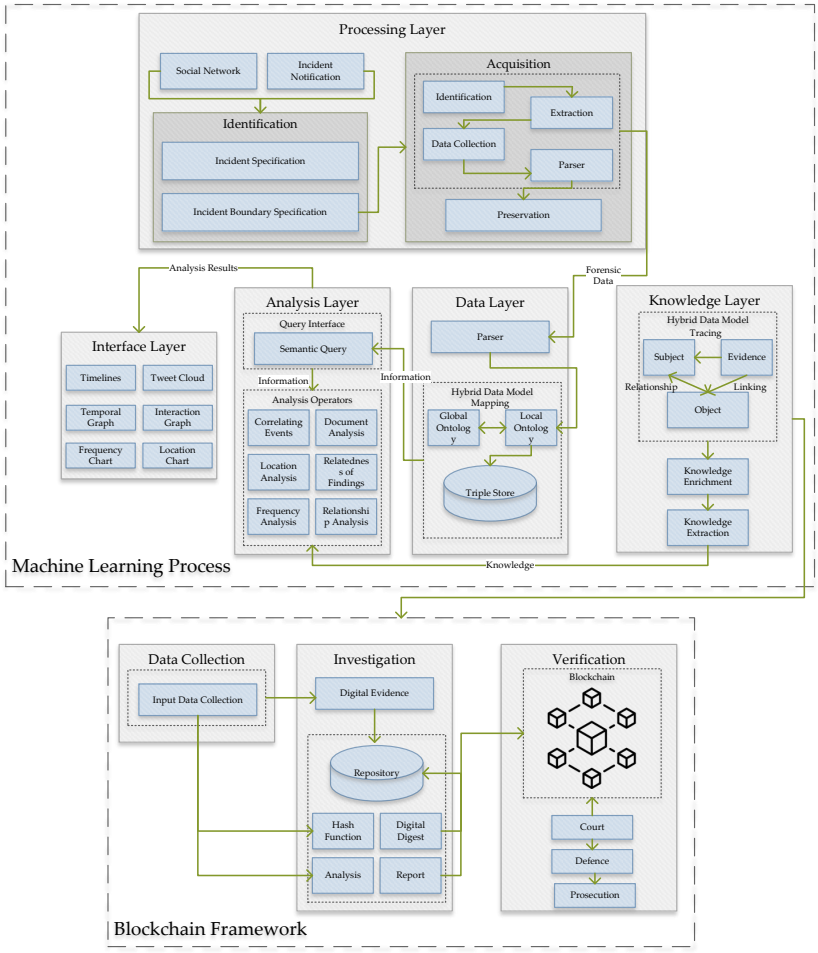

Figure 1. Overview of the proposed digital forensic analysis.

3.1. NLP-Based Digital Forensic Analysis

The presented knowledge model is an event-based system that prepares the social media analysis based on electronic forensics. The ontology technique is applied for representing the related knowledge of OSNs. The detailed explanations show the automated method process and provide formal information through ontologies for the true validation and automated techniques. The investigation of forensic models is from the collection of semi-automated processes. This model contribution provides the boundary identification of data collection from the social media distribution network. The model gives the limitations of forensic data in terms of appropriate parameters and automated collection. Figure 2 describes the details of every layer for the forensic data-analysis process: the automated operators, semantic querying, the rules of the ontology and taxonomy processes, and the identification of the data interchange regarding the defined layers processed for analysis of forensic data system. The data layer contains the parser, the data profile, the content, the data from the network, and activity data. The knowledge layer contains the local/global

ontology process and mapping details. The analysis layer gives the information and further processes the operators, timeline, interaction, temporal patterns, and correlation analysis, and finally, the interface layer shows the user applications and interfaces.

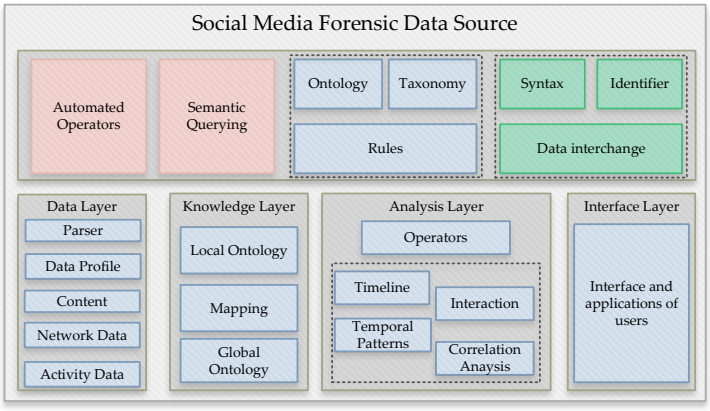

Figure 2. Multi-layered implementation process.

The vectorization process in this approach is based on the latent Dirichlet allocation (LDA) to group some of the topics out of the data. LDA is a famous topic-discovery or topic-categorization approach that clearly separates the content into the clusters of similar data [57]. Each cluster contains similar information and the same direction in terms of meaning and content similarity or probability. Equation (1) presents the estimated topics t based on LDA and edge q_n, which transforms to the vector β_n that provides the probability of $R(z_m|q_n)$ for every topic.

$$\beta_n = (R(z_1|q_n), R(z_2|q_n), \ldots, R(z_m|q_n), \ldots, R(z_t|q_n)) \quad (1)$$

For deciding the topic t, the perplexity model of LDA was applied. This model shows the model's performance and how well it works. Equation (2) shows the process of the perplexity evaluation, where $R(w)$ is the words' probability output from the LDA model, and i is the number of words. Thus, the presented approach evaluates the LDA model perplexity for the vectorization.

$$preplexity = q^{\frac{\Sigma log(R(w))}{i}} \quad (2)$$

Equation (3) presents the vertices β_m, which can be vectorized with a vector in m edges.

$$A_k = (\beta_1, \beta_2, \ldots, \beta_m) \quad (3)$$

Every vertex can have various numbers; in Equation (4), the vertex normalization is evaluated for the topic distribution.

$$A_k = \frac{(\beta_1, \beta_2, \ldots, \beta_m)}{I} \quad (4)$$

Based on the various generated vectors' sizes, the last step is high dimensional. The presented system evaluates the feature relevancy composition to reduce the dimensions. The CFR algorithm is applied for the feature selection regarding the information that can discover topics' degree of impact. Table 2 shows the details of the feature selection of the presented system.

Table 2. Vertices before and after feature selection.

Before Feature Selection						
Samples	topic 1	topic 2	topic 3	topic 4	topic 5	y (label)
Vector 1	1	0	0	0	0	0
Vector 2	0	0	0.7	0.2	0.4	1
Vector 3	0.5	0.3	0.2	0	0.4	0
Vector 4	0	0.6	0	0	0.6	1
After Feature Selection						
Vectors	topic 1	topic 2	topic 3		y (label)	
Vector 1	1	0	0		0	
Vector 2	0	0	0.2		1	
Vector 3	0.5	0.3	0		0	
Vector 4	0	0.6	0		1	

3.2. Blockchain-Based Digital Forensic Analysis

The blockchain approach for the digital forensics process is used to secure the forensic data in terms of transparency and performance. As shown in Figure 3, each entity links together, e.g., users, devices, evidence items, etc. The significant part to guarantee the digital evidence integrity is based on the hierarchy level in an investigation of chains.

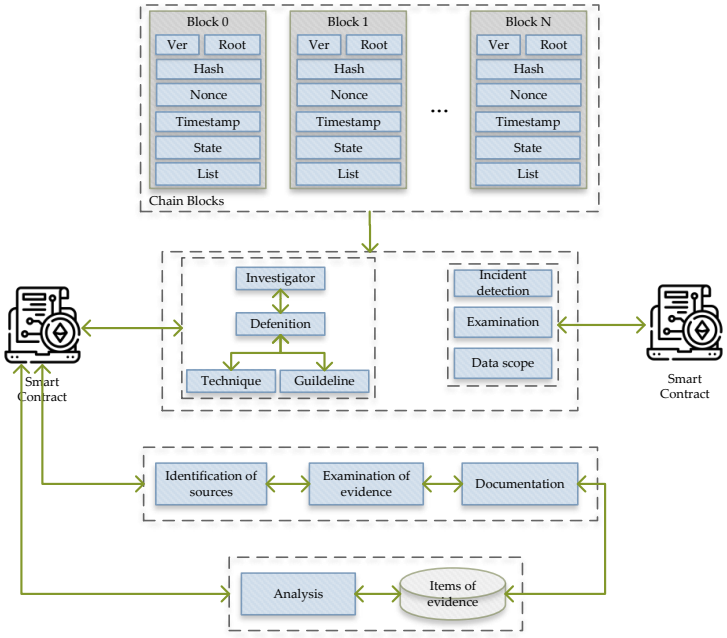

Figure 3. Blockchain-based evidence identification.

There are three main processes defined for the DF investigation: applying a smart contract to perform the evidence analysis automatically, e.g., email analysis, signature or file analysis, etc, and providing better auditability by improving the investigation transparency, thereby reducing the costs and used resources and increasing connection stability between third parties.

4. Experimental Results and Development Environment

This section describes the details of the collected dataset, the system performance evaluation, the experiments, and the results of NLP and blockchain in forensic data analysis. Table 3 shows the details of the development environment for the digital forensic analysis.

Table 3. Development environment.

Module	Component	Description
Machine Learning	Operating System	Microsoft Windows 10
	CPU	Intel (R) Core (TM) i7-8700@3.20 GHz
	Main Memory	16GB RAM
	Core Programming Language	Python
	IDE	PyCharm Professional 2020
	ML Algorithm	Random Forest
Blockchain Framework	Operating System	Ubuntu Linux 18.04 LTS
	Docker Engine	Version 18.06.1-ce
	Docker Composer	Version 1.13.0
	IDE	Composer Playground
	Programming Language	Node.js

4.1. Data Representation and Collection

The data collected in this system is from online social media (OSN), which implements the knowledge model by using ontologies and semantic web processes. The data collection is from famous social media websites, such as Facebook and Twitter, including comments, shares, news broadcasts, etc. The number of users in this environments is high and information sharing is very fast and impressive. Regarding this process, the records of fake shared information is also very high. In this process, we have used 80% of the collected dataset for the training set and 20% for the testing set. Table 4 presents the details of the collected dataset for this approach.

Table 4. Data information.

Data Type	Total Records
Facebook	5000
Twitter	6500
Blogs	6600
News	5500
Training Set	80%
Testing Set	20%

The data layer is responsible for normalizing the provided data and storing them in persistent memory. This memory implements a design that is based on web schema. The unstructured data analysis requires a developed and customized tool for further processing. The analysis layer presents the analysis operators for the automatic process of social media contexts. The computerized analysis method is applied for quick data analysis and evaluation in this process. The decision-making process is a representative evaluation of the human examiner regarding various crimes in the social network evidence. Figure 4 shows the details of data classification and automation solutions.

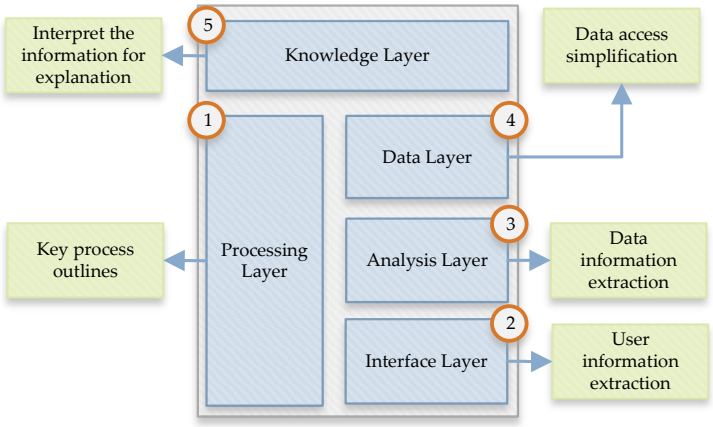

Figure 4. Multi-layered data processing.

Table 5 presents the details of the operators for the data analysis process. Eight operators use the subject and object correlation to analyze the dataset's contents.

Table 5. Analysis operators list of the following processes.

Name of Operators	Details
Tweet Cloud	Object correlation method to provide the fast overview of users' tweet topics.
Hashtag Cloud	Object correlation based on hashtags of user tweets.
Interaction Graph	Subject and object correlation for sorting contacts between the social graph of users with the highest communication frequency.
Interaction Frequency Analysis	Subject and objective correlation to perform the frequency analysis between two users and identify the relationship of the users' communication.
Views Similarity	Rule-based correlation for nearest user-opinion identification.
Trace Operator	Linking the evidence to the entity.
Temporal Activity Graph	Using temporal correlation to analyze the user activity patterns in a defined period.
Geo-location Activity Graph	Object correlation for sorting the location based on the tagged online content.

4.2. Performance Evaluation of the Proposed Online Digital Forensic Analysis

This part presents the performance evaluation of the proposed online forensic analysis. We have defined three metrics of precision P, recall R, and F-measure $F1$. In this process, we used the Random Forest algorithm to analyze this process and compare our results with the Decision Tree, Naive Bayes, Logistic Regression, and Support Vector Machine algorithms. The main reason for using Random Forest in this process is its good performance in terms of classification, as compared to the other algorithms [58]. Table 6 shows the details of each classifier's performance for each fold. Equations (5)–(7) show the details of precision, recall, and f-measure in this process.

$$Precision = \frac{T.Positive}{T.Positive + F.Positive} \quad (5)$$

$$Recall = \frac{T.Positive}{T.Positive + F.Negative} \quad (6)$$

$$Accuracy = \frac{T.Positive + T.Negative}{Total} \quad (7)$$

Figure 5 shows the perplexity records achieved from LDA and records 210 out of the 250 tested topics. Regarding this process, the data was vectorized for the 210 topics. In Figure 5, the x-axis presents the number of topics extracted from this process and the y-axis presents the perplexity of each topic category.

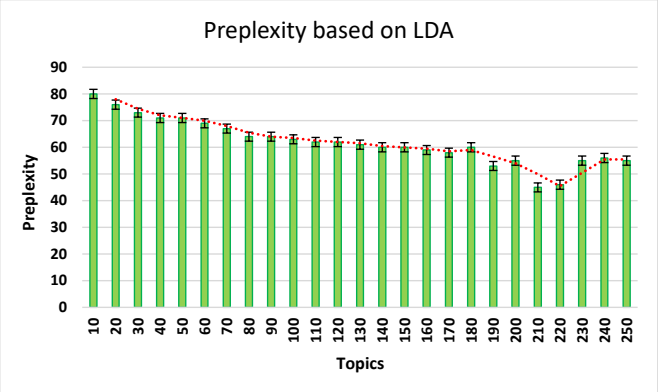

Figure 5. LDA-based preplexity records.

The next step is the cross-validation process. For each five-fold output, the classifier builds the m topics $1 \leq m \leq 210$ in the training set and validates the highest performance record. Table 6 gives the details of defined three classifiers per fold, and Table 7 shows the further evaluations of the metrics' average scores.

Table 6. Different classifiers' performance evaluation for each fold.

Fold	Metrics	Decision Tree	Naive Bayes	Logistic Regression	Random Forest	Support Vector Machine
1	P	0.6595	0.8254	0.9486	0.9846	0.6487
	R	0.7511	0.9700	0.7111	0.6911	0.7348
	F1	0.6667	0.6174	0.8611	0.7811	0.7794
2	P	0.7198	0.3541	0.6736	0.8947	0.4955
	R	0.5511	0.7511	0.4711	0.5948	0.6564
	F1	0.6944	0.5656	0.5611	0.7182	0.5836
3	P	0.6111	0.6993	0.8793	0.9831	0.6939
	R	0.6311	0.9300	0.7511	0.8334	0.8479
	F1	0.6825	0.6111	0.7622	0.7939	0.7749
4	P	0.5968	0.7986	0.8611	0.9444	0.6232
	R	0.8711	0.8711	0.7911	0.7746	0.7498
	F1	0.6929	0.6477	0.7994	0.8337	0.6949
5	P	0.6374	0.4058	0.7929	0.8478	0.5498
	R	0.5111	0.8711	0.7111	0.6964	0.7699
	F1	0.5990	0.6566	0.7633	0.7982	0.6479

Table 7. Average score of different classifiers.

Metrics	Decision Tree	Naive Bayes	Logistic Regression	Random Forest	Support Vector Machine
P	0.6449	0.5367	0.6393	0.0.9443	0.6279
R	0.6631	0.9191	0.6871	0.6943	0.7432
F1	0.6673	0.6197	0.7334	0.7611	0.6745

The presented system shows the benefits of feature selection in this process. Table 8 shows the details of the analysis with and without feature selection. The improvement of Random Forest's performance is very visible. From the perspective of digital investigators, the feature selection is suitable to sort the related topics.

Table 8. Records with and without feature selection.

#	Metrics	Decision Tree	Naive Bayes	Logistic Regression	Random Forest	Support Vector Machine
With feature selection	P	0.6449	0.6367	0.8513	0.9443	0.6279
	R	0.6631	0.9191	0.6871	0.6943	0.7432
	F1	0.6673	0.6197	0.7534	0.7611	0.6745
Without feature selection	P	0.6293	0.6176	0.8122	0.8321	0.4574
	R	0.6171	0.5351	0.6791	0.5467	0.6831
	F1	0.6372	0.5779	0.6974	0.6998	0.5445

The other benefit of applying sorted topics is identifying the communication between networks with a significant volume of data. Figure 6 shows the test set of 106 topics' probability for each fold.

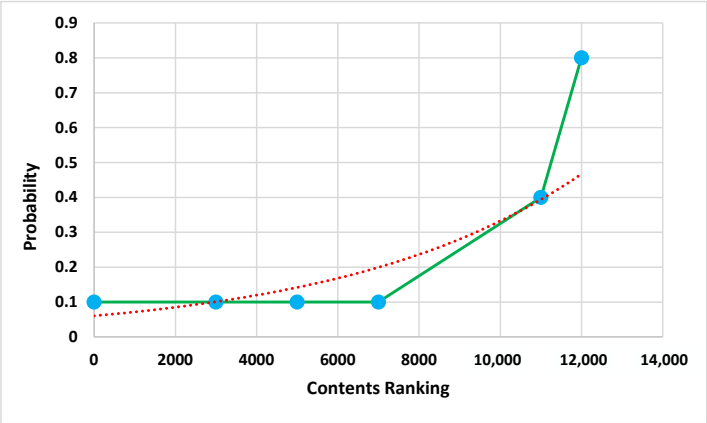

Figure 6. Topic probability analysis records.

4.3. Security Analysis of Online Digital Forensic Based on Blockchain

Security in forensic data is one of the most important and challenging aspects in this area. We have used the blockchain framework for online digital forensic security analysis to improve this system's transparency and rate of the trust according to the following steps:

- The first step is digital evidence identification. The aim of this is to identify the digital fingerprint of evidence. Furthermore, one fingerprint is generated to examine the event for every certain claim;

- Based on the timestamp and additional information, the fingerprint records are written into the evidence block and appended to the blockchain;
- In the blockchain network, every participant holds a copy of the evidence blockchain.

Figure 7 shows the JSON script for the evidence block.


```
Blockchain Header  {8}
    Header Hash : 65e0c3361da2f96c02542a1e798f2
    Version : 03000000
    Previous Block Header Hash : 65e0c3361da2f96c02542a1e798f2
    Merkle Root Hash : 65e0c3361da2f96c02542a1e798f2
    Time : 1516349983
    nBits : 40c42b29
    nonce : fe9f0975
Transactions  {5}
    OpCode : OP_TRUE
    Address Conversion : 123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
    Raw Transaction Format : Information
    CompactSize Unsigned integer : 616
Pending Transactions  {8}
Attachments  {3}
 0  {6}
     content : dGVzdGZpbGU=
     file_name : text/plain
     size : 8
     disposition : attachment
1 {6}
```


Figure 7. Evidence block in JSON script.

Figure 8 presents the details of using blockchain in forensic analysis. There are four main sections, namely, data acquisition, identification, analysis, and presentation. Regarding the timeline, the transactional evidence record is in the blockchain framework. In data acquisition section, all the related information is saved in the blockchain. In the identification section, suspicious files are saved in the blockchain too. In the analysis stage, by using the hash function, various file types are analyzed and stored in blockchain. The presentation stage writes all the reports and findings to the blockchain.

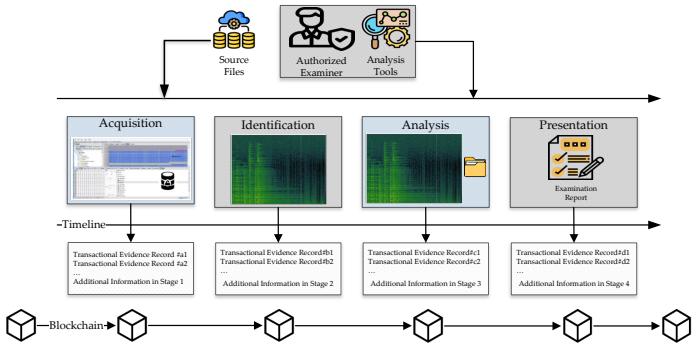

Figure 8. Details of the process of the blockchain framework for forensic analysis.

5. Conclusions

Social media communication is an important source of evidence for criminal investigations, such as fake news or fake election investigations. In this paper, we proposed the integration of NLP techniques with blockchain to improve the security and performance of online digital forensics. In terms of NLP, the LDA topic modeling, feature extraction, and data analysis were applied for a detail analysis of the collected dataset. The collected information is from multi-source social media platforms, which provides more opportunity for results comparisons in various aspects, as compared with other state-of-the-art approaches. The Random Forest algorithm was applied on a real-world dataset and compared with the other four algorithms, namely, the Decision Tree, Naive Bayes, Logistic Regression, and Support Vector Machine algorithms. The main reason for selecting Random Forest for this system is the higher performance of this algorithm in classification tasks and related processes. The concept of blockchain in this system is to improve system security and trace process changes. The defined system is processed in the Hyperledger Fabric framework.

The blockchain framework gives the opportunity to the system process of saving and securing the results, as well as all the digital forensic processess and data, with details. Future studies in this topic can apply the presented system to cybercriminal activities and fraud to overcome the recent issues in this field.

Author Contributions: Data curation, Z.S.; funding acquisition, Y.-C.B.; investigation, Z.S.; methodology, Z.S.; writing—original draft, Z.S.; supervision, Y.-C.B.; project administration Y.-C.B.; validation, Z.S.; visualization, Y.-C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Ministry of SMEs and Startups (MSS), Korea, under the "Startup growth technology development program (R&D, S3125114)", and by the Ministry of Small and Medium-sized Enterprises (SMEs) and Startups (MSS), Korea, under the "Regional Specialized Industry Development Plus Program (R&D, S3246057)", supervised by the Korea Institute for Advancement of Technology (KIAT).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Di Domenico, G.; Sit, J.; Ishizaka, A.; Nunan, D. Fake news, social media and marketing: A systematic review. *J. Bus. Res.* **2021**, *124*, 329–341. [CrossRef]
2. Grubl, T.; Lallie, H.S. Applying Artificial Intelligence for Age Estimation in Digital Forensic Investigations. *arXiv* **2022**, arXiv:2201.03045.
3. Suryanto, H.; Degeng, I.N.S.; Djatmika, E.T.; Kuswandi, D. The effect of creative problem solving with the intervention social skills on the performance of creative tasks. *Creat. Stud.* **2021**, *14*, 323–335. [CrossRef]
4. Shahbazi, Z.; Byun, Y.C. Analyzing the Performance of User Generated Contents in B2B Firms Based on Big Data and Machine Learning. *Soft Comput. Mach. Intell. J.* **2021**, *1*, 1–9.
5. Shahbazi, Z.; Byun, Y.C. Twitter Sentiment Analysis Using Natural Language Processing and Machine Learning Techniques. In Proceedings of the KIIT Conference, Jeju, Korea, 15 December 2021; pp. 42–44.
6. Shahbazi, Z.; Byun, Y.C. Deep Learning Method to Estimate the Focus Time of Paragraph. *Int. J. Mach. Learn. Comput.* **2020**, *10*, 75–80. [CrossRef]
7. Heckmann, T.; Souvignet, T.; Sauveron, D.; Naccache, D. Medical Equipment Used for Forensic Data Extraction: A low-cost solution for forensic laboratories not provided with expensive diagnostic or advanced repair equipment. *Forensic Sci. Int. Digit. Investig.* **2021**, *36*, 301092. [CrossRef]
8. Patil, A.; Banerjee, S.; Jadhav, D.; Borkar, G. Roadmap of Digital Forensics Investigation Process with Discovery of Tools. *Cyber Secur. Digit. Forensics* **2022**, 241–269.
9. Rouzbahani, H.M.; Dehghantanha, A.; Choo, K.K.R. Big Data Analytics and Forensics: An Overview. In *Handbook of Big Data Analytics and Forensics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–5.
10. Li, S.; Sun, Q.; Xu, X. Forensic analysis of digital images over smart devices and online social networks. In Proceedings of the 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 1015–1021.
11. Javed, A.R.; Ahmed, W.; Alazab, M.; Jalil, Z.; Kifayat, K.; Gadekallu, T.R. A Comprehensive Survey on Computer Forensics: State-of-the-art, Tools, Techniques, Challenges, and Future Directions. *IEEE Access* **2022**, *10*, 11065–11089. [CrossRef]
12. Lorch, B.; Scheler, N.; Riess, C. Compliance Challenges in Forensic Image Analysis Under the Artificial Intelligence Act. *arXiv* **2022**, arXiv:2203.00469.
13. Hemdan, E.E.D.; Manjaiah, D. An efficient digital forensic model for cybercrimes investigation in cloud computing. *Multimed. Tools Appl.* **2021**, *80*, 14255–14282. [CrossRef]
14. Alnajjar, I.A.; Mahmuddin, M. Feature indexing and search optimization for enhancing the forensic analysis of mobile cloud environment. *Inf. Secur. J. Glob. Perspect.* **2021**, *30*, 235–256. [CrossRef]
15. Bhagat, S.P.; Meshram, B.B. Digital Forensic Tools for Cloud Computing Environment. In *ICT with Intelligent Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 49–57.
16. Salamh, F.E.; Karabiyik, U.; Rogers, M.K.; Matson, E.T. A comparative uav forensic analysis: Static and live digital evidence traceability challenges. *Drones* **2021**, *5*, 42. [CrossRef]

17. Khalid Alabdulsalam, S.; Duong, T.Q.; Raymond Choo, K.K.; Le-Khac, N.A. An efficient IoT forensic approach for the evidence acquisition and analysis based on network link. *Log. J. IGPL* **2022**. [CrossRef]
18. Loli, M.; Mitoulis, S.A.; Tsatsis, A.; Manousakis, J.; Kourkoulis, R.; Zekkos, D. Flood characterization based on forensic analysis of bridge collapse using UAV reconnaissance and CFD simulations. *Sci. Total Environ.* **2022**, *822*, 153661. [CrossRef] [PubMed]
19. Li, S.; Qin, T.; Min, G. Blockchain-based digital forensics investigation framework in the internet of things and social systems. *IEEE Trans. Comput. Soc. Syst.* **2019**, *6*, 1433–1441. [CrossRef]
20. Alsulami, H. Implementation analysis of reliable unmanned aerial vehicles models for security against cyber-crimes: Attacks, tracebacks, forensics and solutions. *Comput. Electr. Eng.* **2022**, *100*, 107870. [CrossRef]
21. Misra, S.; Arumugam, C. *Illumination of Artificial Intelligence in Cybersecurity and Forensics*; Springer: Berlin/Heidelberg, Germany, 2022; Volume 109.
22. Kaushik, K.; Dahiya, S.; Sharma, R. Role of Blockchain Technology in Digital Forensics. In *Blockchain Technology*; CRC Press: Boca Raton, FL, USA, 2022; pp. 235–246.
23. Kebande, V.R.; Ikuesan, R.A.; Karie, N.M. Review of Blockchain Forensics Challenges. In *Blockchain Security in Cloud Computing*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 33–50.
24. Li, S.; Choo, K.K.R.; Sun, Q.; Buchanan, W.J.; Cao, J. IoT forensics: Amazon echo as a use case. *IEEE Internet Things J.* **2019**, *6*, 6487–6497. [CrossRef]
25. Li, S.; Zhao, S.; Yang, P.; Andriotis, P.; Xu, L.; Sun, Q. Distributed consensus algorithm for events detection in cyber-physical systems. *IEEE Internet Things J.* **2019**, *6*, 2299–2308. [CrossRef]
26. Ganesh, N.; Venkatesh, N.; Prasad, D. A Systematic Literature Review on Forensics in Cloud, IoT, AI & Blockchain. *Illum. Artif. Intell. Cybersecur. Forensics* **2022**, *109*, 197–229.
27. Rajawat, A.S.; Rawat, R.; Barhanpurkar, K. Security Improvement Technique for Distributed Control System (DCS) and Supervisory Control-Data Acquisition (SCADA) Using Blockchain at Dark Web Platform. *Cyber Secur. Digit. Forensics* **2022**, 317–333. [CrossRef]
28. Shahbazi, Z.; Byun, Y.C. Blockchain-based Event Detection and Trust Verification Using Natural Language Processing and Machine Learning. *IEEE Access* **2021**, *10*, 5790–5800. [CrossRef]
29. Ryu, J.H.; Sharma, P.K.; Jo, J.H.; Park, J.H. A blockchain-based decentralized efficient investigation framework for IoT digital forensics. *J. Supercomput.* **2019**, *75*, 4372–4387. [CrossRef]
30. Siddiqi, A.S.; Alam, M.; Mehta, D.; Zafar, S. Machine Learning-Based Predictive Analysis to Abet Climatic Change Preparedness. In *Cyber Security and Digital Forensics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 541–550.
31. Mishra, A.; Khan, M.; Khan, W.; Khan, M.Z.; Srivastava, N.K. A Comparative Study on Data Mining Approach Using Machine Learning Techniques: Prediction Perspective. In *Pervasive Healthcare*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 153–165.
32. Shahbazi, Z.; Byun, Y.C. Fake media detection based on natural language processing and blockchain approaches. *IEEE Access* **2021**, *9*, 128442–128453. [CrossRef]
33. Coffey, C.A.; Batastini, A.B.; Vitacco, M.J. Clues from the digital world: A survey of clinicians' reliance on social media as collateral data in forensic evaluations. *Prof. Psychol. Res. Pract.* **2018**, *49*, 345. [CrossRef]
34. Baror, S.O.; Venter, H.S.; Adeyemi, R. A natural human language framework for digital forensic readiness in the public cloud. *Aust. J. Forensic Sci.* **2021**, *53*, 566–591. [CrossRef]
35. Barik, K.; Abirami, A.; Konar, K.; Das, S. Research Perspective on Digital Forensic Tools and Investigation Process. In *Illumination of Artificial Intelligence in Cybersecurity and Forensics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 71–95.
36. Shahbazi, Z.; Byun, Y.C. Agent-Based Recommendation in E-Learning Environment Using Knowledge Discovery and Machine Learning Approaches. *Mathematics* **2022**, *10*, 1192. [CrossRef]
37. Kaur, R.; Singh, S.; Kumar, H. Authorship analysis of online social media content. In Proceedings of the 2nd International Conference on Communication, Computing and Networking, Haldia, India, 15–16 November 2019; pp. 539–549.
38. Montasari, R.; Hill, R. Next-generation digital forensics: Challenges and future paradigms. In Proceedings of the 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3), London, UK, 16–18 January 2019; pp. 205–212.
39. McGuire, J.C.; Leung, W.S. Enhancing digital forensic investigations into emails through sentiment analysis. In Proceedings of the ECCWS 2018 17th European Conference on Cyber Warfare and Security V2, Oslo, Norway, 28–29 June 2018; p. 288.
40. Shahbazi, Z.; Byun, Y.C. Computing focus time of paragraph using deep learning. In Proceedings of the 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), Seogwipo, Korea, 8–10 May 2019; pp. 1–4.
41. Shahbazi, Z.; Byun, Y.C. LDA Topic Generalization on Museum Collections. In *Smart Technologies in Data Science and Communication*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 91–98.
42. Mouhssine, E.; Khalid, C. Social big data mining framework for extremist content detection in social networks. In Proceedings of the 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Rabat, Morocco, 21–23 November 2018; pp. 1–5.
43. Dhaliwal, P. Comprehensive Exploration of Machine Learning based models in Digital Forensics—A plunge into Hate Speech Detection. In Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 17–18 December 2021; pp. 1933–1938.
44. Iqbal, F.; Debbabi, M.; Fung, B. Artificial intelligence and digital forensics. In *Machine Learning for Authorship Attribution and Cyber Forensics*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 139–150.

45. Horan, C.; Saiedian, H. Cyber Crime Investigation: Landscape, Challenges, and Future Research Directions. *J. Cybersecur. Priv.* **2021**, *1*, 580–596. [CrossRef]
46. Shahbazi, Z.; Byun, Y.C.; Lee, D.C. Toward representing automatic knowledge discovery from social media contents based on document classification. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 14089–14096.
47. Shahbazi, Z.; Byun, Y.C. Topic prediction and knowledge discovery based on integrated topic modeling and deep neural networks approaches. *J. Intell. Fuzzy Syst.* **2021**, 41, 1–17. [CrossRef]
48. Seckiner, D.; Mallett, X.; Roux, C.; Meuwly, D.; Maynard, P. Forensic image analysis—CCTV distortion and artefacts. *Forensic Sci. Int.* **2018**, *285*, 77–85. [CrossRef]
49. Sanyasi, M.; Kumar, P. Digital Forensics Investigation for Attacks on Artificial Intelligence. *SPAST Abstr.* **2021**, *1*, 1.
50. Khan, A.A.; Shaikh, A.A.; Laghari, A.A.; Dootio, M.A.; Rind, M.M.; Awan, S.A. Digital forensics and cyber forensics investigation: security challenges, limitations, open issues, and future direction. *Int. J. Electron. Secur. Digit. Forensics* **2022**, *14*, 124–150. [CrossRef]
51. Krishnan, S.; Shashidhar, N.; Varol, C.; Islam, A.R. Evidence Data Preprocessing for Forensic and Legal Analytics. *Int. J. Comput. Linguist. (IJCL)* **2021**, *12*, 24.
52. Choi, J.; Yu, J.; Hyun, S.; Kim, H. Digital forensic analysis of encrypted database files in instant messaging applications on Windows operating systems: Case study with KakaoTalk, NateOn and QQ messenger. *Digit. Investig.* **2019**, *28*, S50–S59. [CrossRef]
53. Zhang, H.; Chen, L.; Liu, Q. Digital forensic analysis of instant messaging applications on android smartphones. In Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC), Maui, HI, USA, 5–8 March 2018; pp. 647–651.
54. Du, X.; Hargreaves, C.; Sheppard, J.; Anda, F.; Sayakkara, A.; Le-Khac, N.A.; Scanlon, M. SoK: Exploring the state of the art and the future potential of artificial intelligence in digital forensic investigation. In Proceedings of the 15th International Conference on Availability, Reliability and Security, Virtual Event, Ireland, 25–28 August 2020; pp. 1–10.
55. Xiao, J.; Li, S.; Xu, Q. Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access* **2019**, *7*, 55432–55442. [CrossRef]
56. Al-Jadir, I.; Wong, K.W.; Fung, C.C.; Xie, H. Enhancing digital forensic analysis using memetic algorithm feature selection method for document clustering. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 3673–3678.
57. Venugopalan, M.; Gupta, D. An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. *Knowl.-Based Syst.* **2022**, *246*, 108668. [CrossRef]
58. Palimkar, P.; Shaw, R.N.; Ghosh, A. Machine learning technique to prognosis diabetes disease: Random forest classifier approach. In *Advanced Computing and Intelligent Technologies*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 219–244.