Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 58-72
Received 15 April 2021; Accepted 05 May 2021.

A Review on Data Privacy Detection in Social Networks using Data Mining, Machine Learning and Blockchain technologies

P.S.Arun kaarthi[1], S.Sathiyabama[2]
[1]Research Scholar, Department of Computer Science, Periyar University, Salem, Tamilnadu, India.
[2]Assistant Professor, Department of Computer Science, Thiruvalluvar Government Arts College, Rasipuram, Tamilnadu, India.

Abstract.

Data Privacy Detection in Social Networks (DPDSNs) is overriding the need for the privatization of information in online networks and social media. In recent times, data security and privatization have become the biggest challenge for all categories of people ranging from affluent people to familiar people in society. Social networking data being sensitive is capable of causing a big difference in all the domains where sensitive information cases sensations in positive as well as in adverse impacts. The primary objective of this paper is to conduct a deep- rooted review on various aspects of Data Privacy Detection methods applied in Social Networks using combinational analysis of multiple areas like Data Mining, Machine Learning, and Blockchain Technological processes respectively. Numerous papers were surveyed and reviewed with a primary focus on methodologies used, techniques applied, algorithms designed as well as the outcomes of results achieved in the course of the study. After consequent reviews and analysis, a novel framework based on the non-existence of Data Privacy Detection Techniques was proposed, and the future directions are shared to counteract the problems faced in the data privacy along with this study.

Keywords: Data Privacy Detection in Social Networks (DPDSNs), Data Mining, Machine Learning Techniques, Blockchain Technology, Social Networks, Sensitive data

INTRODUCTION

The world being an information-based platform operates on the data available and extracts the knowledge based on the data using Data Mining techniques. The prediction and detection methods are applied based on the Machine Learning Algorithms as they are capable of designing algorithms to predict the same using artificial intelligence. Hence the DPDSNs combine the detection and prediction methods to identify the best factors that would be handled to detect the privacy and protect the social networking information from intruders and information stealers. Various techniques and methods were created, new frameworks and models were proposed to assess data security and privacy. Hence the primary objective of the paper is to conduct a deep-rooted survey on how various problems in data privacy have been addressed using three powerful techniques viz Data Mining, Machine Learning, and Blockchain Technology, respectively. The general Social media privacy problems were reviewed in the earlier part of the study. Also, the Data Mining assessments, Machine
Learning Models used and the impact of Blockchain Technology in protecting Data Privacy problems has been examined in the overall context of the study.

LITERATURE REVIEW

Social Networks

Social Networking denotes the mode to connect the people and also a means of communication among them. Social Networking has developed itself as a mass media which is capable of turning the heads of any personalities and also can bring changes in the political, social, and economic condition of the country. Hence data privacy in social networking is considered highly sensitive and requires a robust framework to protect sensitive information from intruders. Various researchers have performed further analysis and predictions on social networking in different research works. Analytics on Delhi Odd-even Policy based on the tweets posted on Twitter during the period from December 2015 to August 2016. This research analysis converted unstructured tweets into structured information based on open source libraries. The tweets collected was based on hashtags API on which sentiment predictive models are proposed. The performance evaluation was conducted, and the results showed that TextBlob API, along with the proposed model, overwhelmed the other four sentimental models used earlier [1].

A novel model framework was developed that combined finds the solution for the link and interaction polarity problems and predicts the results thereof. The experiments concluded that the proposed model framework was sufficient to assist data sparsity as well as cold start problems linked with social networking links. The paper was classified under link mining concepts[2]. The Social networking usage of clickable buttons was tested based on the polarised public opinion that gave a response for corporate social advocacy. The major participants were the advocators and boycotters of president Trump's immigration ban in 2017. The assessment resulted in the identification of three unique nature of boycotter in social networking as well as advocators. However, the boycotters were very dense compared to advocators. The results could impact the decision made by the president and create an impact on the continuation of the ban[3].

An existing social network analysis was performed on Facebook and converted the data into knowledge. New patterns were developed through the aggregation of various interactions among online users on Facebook on disaster happenings[4]. The social network analysis focused on disaster news spread frequency and information propagation in an emergent manner. An exploration of the various opportunities in avoiding money laundering was studied using the central database operated in Italy. A novel model was proposed to sort out and map the relational data among the information from social networks. The collected data was developed as a predictive model to assess financial operations and network transactions from various countries. The research study finally developed patterns that could identify the potential intruders and criminals in social networking in the form of clusters[5].

A conflict Relationship Investigation process was introduced based on social network analysis where the analysis is based on sparse representations. The principal objective was to solve LSDM problems using the decision selection process as well as DM's Weights respectively. The proposed model S-CRIP was identified as an up-gradation of various other models like the CD-CRIP model that can handle numerical representations. The model was very highly influential in building numerical analysis in future social network analysis where text-based analysis is possible[6]. A method to detect the repetitive, abusive languages of the social network users was detected using the development of a cyberbullying phenomenon monitor and also to combine classification of messages with social network analysis. The classification module was evaluated and was built on the inputs received through Instagram messages. Finally, the user interface for detecting cyberbullying was designed and was successful[7].

A novel framework was developed to secure and store social network data in a cloud-based environment. The framework was designed in such a way that it encrypts the original data and stores it in the cloud for further processing[8]. Whenever the data is accepted using a private key from the right user, the contents are decrypted to acquire the original data and view the contents for the social network user in connection. The Situation deliberately makes communication in social networking more reliable and avoids leakage of messages and other secure information over Social networks. A comprehensive review of the different security and protective issues faced by social networking media was conducted[9]. The study focused on the various security methods portrayed in all the social networking components and also gives the security measures sufficient to suppress the threats. The review also identified strategies to make the interactions of social networks more promising shielded with prevention from unexpected data or intrusions. The study was beneficial to identify the flaws in social networking with regards to data privacy. Various papers related concepts were studied on Online Social Networks (OSNs) to identify the methods used to classify the original and Fake profiles of various Social Networks. The security issues and privacy problems were studied along with their relevance to the information shared during communication. Different other analysis was also carried out based on the problems in security, modeling tools, and even on other security threats of social networks in online mode[10]. A Forensic Analysis on Instagram was carried out based on social networks. Forensic techniques are applied to pictures or posts on Instagram to detect the various parameters like comments, likes, and tags to find the intruders or fake profiles in social networks. Social networking forensics enables us to find the traces left by the intruders after unsolicited messages or known intrusions. This research gave proof to test the user behaviors and also assisted on the system based forensic analysis to find pieces of evidence on any privacy intrusion in social network accounts of legitimate users[11].

A similar analysis was conducted on the different watermarking techniques applied in social networks using a novel model called DWT, DCT, and SVD based methods that decompose the image lowers the frequency bands embeds the image to find the information related to the idea like the owner, health details, etc. that can be applied in various fields[12]. The newly formed model created instantaneous methods to find the quality of the image and also the intrusion happened in the social network with the fake identities of the image, respectively.

After the baseline analysis of various social networking problems associated with data privacy and reliability, a review of the techniques to be applied was mandatory. Hence Data Mining techniques were reviewed and analyzed.

Data Mining

Data Mining techniques are one of the chiefly utilized methods for social network analysis ranging from text-based analysis to numerical analysis in all prospects. Hence the assessment of DPDSNS was also proposed based on Data mining techniques. Among the reviews collected[13], performed the survey on PPDM practices and relevant literature to assess the metrics used to evaluate the privacy detection methods and its relevant applications. This model helped to preserve privacy and enhance the storage of sensitive information on social media and the internet.

A framework was developed for providing support for healthcare systems through privacy- preserving process mining techniques[14]. The sensitive information in healthcare sectors was preserved by applying layers of data transformation methods as well as process mining toidentify anonymous healthcare information. A Bayesian-based PPDM method was developed based on classification techniques[15]. The author surveyed various models presented for ten years on privacy-preserving data mining using the data perturbation method, which is found to be independent of the algorithm. The author used practical techniques that are very casual to identify privacy preservation. The results showed that the current method was convincing compared to the existing data privacy detection methods.

Various data mining techniques were applied to detect privacy, threats in security, and also the investigations related to their usage and limitations. The review channelized various problems for data privacy like security threats, privacy protection, denial of service in standalone and distributed systems, botnet, malware, spyware as well as ransomware and probing. The review was presented to assess the future needs of data privacy methods and the enhancements possible[16]. A Data Mining based privacy protection method was developed[17] and used for predicting the length of stay of patients thereby managing the privacy data and coding for management of admissions in healthcare and insurance systems.

A heuristic approach was developed that worked on examining experience as a factor to manage the privacy of visitors in stores. The work included classification techniques in data mining using an algorithm and detected the problems using facts and patterns[18]. The research work is a mixture of theoretical and practical methods and hence gained importance in predicting the privacy of information over the internet. An assessment of Privacy- Preserving Data Mining algorithms was conducted[19] based on various frameworks, patterns created, and also on the usage and tactics followed by various research works in existence. The data privacy framework analysis showed that privateness was in problem based on a few parameters like efficiency, degree of uncertainty, the utility of data as well as resistance to being hidden respectively. The research also confirmed that no data mining- based privacy-preserving algorithm is in existence at present, and maximum frameworks are existing in a theoretical manner only.

A framework for transparency and enhancing the confidence of data privacy in the educational sector was developed[20]. The research study-initiated checklist-based assessments to check for student information transferred over different networks and also dealt with the prediction of loss of privacy of information in different situations based on student dataset. The results are confirmed based on the missing word in the dataset at the end of the transfer of information in online networks.Six data mining methods were tested through the cross-industry standard process for data mining in predicting the bloodstream infections with central line association among patients in healthcare streams. The dataset from the US National Healthcare Safety Network is used for the examination of the loss of sensitive information using data mining algorithms. The AdaBoost algorithms produced the Accuracy of 89.7%, which gave hope for high accuracy predictions in the future[21]. A survey on Knowledge Discovery and Data Mining (KDDM) process were conducted and its impact on handling data privacy and protection mechanisms in South African bank sectors based on user segmentation techniques. The financial preferences and data modulations, fluctuations based on online banking are observed in the survey[22].

A model was developed using data mining algorithms to apply in privacy information among the medical domain. The framework model was created using patterns generated by Euclidean Distance, spectral clustering, etc. Privacy protection is a measure based on loss of data on the sensible part of the data in a dataset[23]. The analysis was based on the different strategies used in the protection of private data through the design of the best tool using clustering and classification tool in Data Mining. The results on the Prima Indians Diabetes dataset showed a feasible Random Index for clustering with 0.52, whereas Parkinson's speech database varied from 0.45 to 0.50, respectively. A novel knowledge-based design for data privacy was introduced[24] which used agile manufacturing techniques with data mining. The product emergence process (PEP) was used to assess the data privacy mining methods in existence. This model also proved to be useful in the context of data mining.

The prediction of the strength of data was demonstrated in social networks using data mining methods like decision tree, naïve Bayesian techniques. Also, ensemble methods of classification were used. The sample analysis was made with LinkedIn data using Bagging and Boosting techniques[25]. Also, performance techniques like Accuracy, F-Measure, and average executing time are proposed as the best evaluation methods for solving the problem. System-based protection for the privacy of data was offered and developed[26] through data mining, three-level hierarchical architecture. The in-depth analysis was carried out based on the data loss in many websites because of the high efficiency and functioning of the architectural system. A novel algorithm PBIDOT was proposed[27] to optimize the results of the prediction of data privacy methods. The novel algorithm was examined for various performance measures like efficiency, scalability accuracy, and resistance of attack. The research outcomes showed promising results with speed of execution, the resistance of attack, Accuracy as well as scalability concerning the protection of privacy of information.

The techniques reviewed based on data mining methods and techniques, the novel frameworks and models, were assessed. The reviews suggested that data mining is the central core to create a hybrid framework that could be supported further with machine learning techniques.

Machine Learning Techniques

The impact of Data Mining on Data Privacy Detection in Social Networks accounted for the basic design methods. However, the detection required complex algorithms and practices that combine the features to enhance the accuracy of prediction. Hence Machine Learning Techniques (MLT) are reviewed to acknowledge the best techniques that can provide a solution for the problem. A study on the use of Machine Learning Techniques would suggest the best methods for the formation of frameworks to detect Data Privacy[28]. The various problems and challenges were surveyed encountered in ERP systems to preserve the privacy of data stored and processed in the system using machine learning techniques. The online data based on the cloud and computational capabilities are held using a modeled dataset as a training set that was capable of predicting privacy issues in any situation. The primary concern is based on the ERP systems used in enterprise organizations and also on how the challenges can be reduced through various machine learning techniques. The research proved that Machine learning techniques could be applied to detect the privacy of information.

Reviewed all the privacy-based problems in federated learning systems. The review was channelized based on six facets viz machine learning model, distribution of data, privacy mechanism, the architecture of communication, federation scale, and federation motivation. Various case studies were also studied and presented in the paper[29]. The well-organized privacy-preserving machine learning scheme was developed for handling multiple providers of data from different sources. This model enables all participants of information to directly verify the shared data using Unidirectional Proxy Re-Encryption (UPRE) technique. The proposed model was able to reduce the costs associated with computation and also relieved the problems related to communication between users[30].

Proposed Sherpa.ai Framework for Federated Learning using Federated learning methods as well as privacy in different situations. The paper gave different ways to identify privacy problems using classification and regression techniques of machine learning models[31].

Examined the leakages that happened in machine learning trained data models to identify and remove the black box reference attacks. The method developed a mechanism that could train models which has the privacy of members. Also, the min-max model was used to find the best-suited model for assessing risk-free regulariser in training neural networks with deep learning applied on benchmark datasets. The model was found to be effective with less loss of privacy and high Accuracy[32].

Examined and presented on the pertinent issue of lack of privacy protection through proposing a machine learning-based model. The method was risk-free and used European Union and General Data Protection Regulation for assessment of data. The results summarised the entire works and also suggested the importance of informal methods and decisions to preserve confidential data[33]. A model was presented to detect physical hardware attacks on systems leading to social networking like denial of service, spoofing, eavesdropping methods, and jamming respectively. IoT based system was designed using machine learning model algorithms and also the data privacy was established in the system using detection of malware as well as access control, authentication, and secure offloading techniques[34].

A novel model was developed that assisted in protecting the datasets from cloud infrastructure from various users using public-key encryption with the double encryption algorithm (DD-PKE)[35]. The model designed was more secure and gave high security compared to other machine learning models. A model was presented that assessed the privacy of sensitive information in a society based on online communication using social networks. It also portrayed various deep learning and machine learning methods to be used along with cryptographic methods[36]. A similar analysis with a survey was conducted[37] on techniques used in privacy preservation based on deep learning and machine learning techniques. The comparative analysis was also made on the existing best practices of PPDL and showed information related to MLaaS environments as well.

Hence the MLT empowers a platform that can combine the best characteristics and algorithms of Data Mining with being incorporated with MLT to form a framework architecture that can enhance the predictions of data at the maximum level possible.

Blockchain Technology

Blockchain Technology qualifies storage of information in the form of blocks of data that are easily differentiable, storage of data under high space, especially social network-based data. The blockchain was identified through reviews as one of the essential techniques that were capable of managing the privacy of information in social networks. A blockchain-based framework model to secure the data, interoperability, as well as the efficiency of medical data. The sensitive information in the healthcare systems was protected using an Ethereum- based blockchain method that prevents the access control of intruders[38]. The communication of Ancile also integrates a framework with privacy protection and detection for a long time. A similar security framework was[39] which assisted in the integration and management of physical, social, and business aspects to increase the efficiency of secure communication in a smart city. Large-scale information architecture was developed using blockchain technology to manage Electronic health records from losing the privacy of information. This architecture is also integrated and found to be useful in all systems[40]. A similar analysis of E-health systems was assessed using blockchain technology[41]. The problems and benefits are evaluated in the study and the performance determined for the same.

A novel data privacy management framework was developed with the aid of blockchain technology to manage the financial records using three mechanisms viz. Data privacy classification method, a collaborative filtering model, and data disclosure confirmation scheme, respectively[42]. A privacy-preserving framework was created to detect the ownership of data, transparency of information as well as user suitability among all the users connected in an online network[43]. The system models were classified based on the owner and the consumer process. The research showed that privacy-preserving is possible with blockchain technology.

The smart home-based IOT blockchain was designed[44] through the creation of scenarios like Ganache, Web3.js, and Remix that was built on blockchain servers. The primary objective was to enhance privacy, develop access control, and also to extend the ability to preserve the data. Various constraints, like data privacy, token usage, policy updations, contracts, and judging the misbehavior, are assessed for data privacy during the study. Studied and compared theoretical formulation and practical methods of data privacy techniques based on trusted third parties, cryptographic procedures, hardware with security as well as blockchain-based practices[45]. The evaluation was carried out in biomedical studies, which resulted in the design of the best paradigms. The blockchain smart contracts are utilized for the system, and it resulted in insufficiency to establish data privacy and hence hybrid methods are recommended for future works. Analyzed the importance of Blockchain technology in smart Healthcare systems through dictation and serving methods. It was studied in the research that multiple persons share the same data resulting in loss of privacy which can be covered through regulation, monitoring as well as sensing the paradigm using blockchain technology4[6.] A user-centric method called ADvoCATE was developed which combined IoT systems to design hardware and software architecture with the aid of blockchain technology that is incorporated using intelligent services[47].

A massive survey on blockchain technology-based upon data privacy detection and preservation methods was conducted[48] that analyzed various algorithms that are consensus in nature. The survey also analyzed multiple trends associated with different periods of blockchain technology that can support data privacy and build a robust framework. Utilized the recently created PoW based protocol called GHOSTDAG that are used to oversimplify the factors like the acyclic graph, scalability as well as throughput of the system. The private blockchains are used to develop smart contracts in handling the health information of patients. The work resolved the privacy and security problems that existed in remote systems[49].

The typical discussion on various algorithms used along with blockchain technologies to prevent data privacy was analyzed[50] which concentrated on smart contracts with their applications for the future. Designed benchmark sensor-based data management and transferring software to preserve clinical data [51]. A review was conducted[52] based on the preservation of healthcare data and also enhanced the security, privacy, control, and storage information of healthcare records. The review also suggested that a framework was required to apply blockchain technologies in data privacy detections.

The reviews suggested that blockchain technology can be applied in hybridization with data mining and machine learning techniques to design a data privacy detection system architecture.

DATA PRIVACY DETECTION IN SOCIAL NETWORKS FRAMEWORKS

DPDSNs enable privacy protections, leakages, and other concerns to protect data as the basal view to review and find the best models that are found suitable for future works. A schematic review was conducted based on the existing models or frameworks in existence that are used in preserving the privacy of data. The overall review analysis was summarised in Table 1.

Table-1: Review of Frameworks in existence related to Data Privacy Detection in Social Networks (DPDSNs)

Materials and Methods	Framework used	Outcomes
To assess social networks online and identify their community networks to preserve sensitive data. The model used connections of social networks and published content of users for evaluation[53].	Community Detection Framework	The Proposed framework outperformed existing baseline algorithms in terms of high accuracy.
Concentrated on privacy- preserving social network conversion methods to identify and protect the privacy of user's data[54]	A faction of Data in Social Network	The data security was inactive in many cases with a low average local accessing frequency of data.
To develop a framework with three levels detecting sensitive information on the social network based on statistical and standard knowledge[55]	Semantics-based sensitive topic diffusion detection framework	Results include identification of sensitive clusters, eliminate nonsensitive terms, Twitter analysis based on Kullback– Leibler (KL) divergence, and high correlation of topics of sensitiveness with 4500 posts of 790 users.
To develop a taxonomy guided multi-task learning model to predict privacy leakage tested among 400 users in a personal network[56]	Personal Privacy-Preserving Framework	Found sequential leakages based on LIWC and Sentence2Vector features. It was also observed with a permanent gap between real and fake audience in the social network.
To detect the activities of	Efficient Activity	Validations on benchmark

humans under smart homes using spatiotemporal mining techniques. Also, a novel framework for predicting the privacy of data is proposed[57]	Recognition Framework	datasets indicated that the accuracy and utility of privacy were high compared to existing models.
To utilize the existing model and incorporate end-user privacy as well as efficiency in detecting violations in the social network[58].	APRIGUARD: The Automatic and Efficient Method for Privacy Violations Detection	The model was capable of preventing data theft before it happens. Also, the Accuracy was 89.5% which outperformed the existing model with 85.1% accuracy.
To develop a framework for preventing spams in a decentralized network on DOSNs[59].	DLSAS: Distributed Large-Scale Anti-Spam Framework	The results after examining Twitter datasets showed high efficiency, robustness with the newly formed framework.
To create a framework that encrypts image contents changes the HSV color and adds a secret code using an algorithm to enhance the privacy of image data[60].	iterative magic matrix encryption algorithm (IMMEA)	The results related to both qualitative and quantitative methods are convincing and found efficient compared to existing models.
To present the novel data privacy framework model using existing software Vigi4Med and implement it in social network analysis to prevent leaking of privacy information[61].	Vigi4Med Scraper	More than 20 websites are analyzed which accounted for 200 GB of Data and found that privacy leakage could be determined using the model.

The reviews indicated a comprehensive idea of creating a hybrid model for enhancing privacy detection as well as the security of data over social networks. The data leakage was also concentrated by many of the authors to protect the privacy of data. It is important to review and analyze the various factors that have to be tested for privacy predictions.

ASSESSMENT FACTORS FOR DATA PRIVACY DETECTIONS IN SOCIAL NETWORKS

The Data privacy factors have to be identified to assess and qualify the performance of Data Privacy Detections in Social Networks. Hence some of the reviews were made and studied based on the existing benchmark models and frameworks. Examined various privacy protection factors that would enable smooth functioning of detecting privacy violations in mobile social networks and online social networks respectively[62]. The survey concentrated on the following factors:

1. Owner Privacy Factors
2. Copyrights
3. Property rights
4. Verified users
5. Confidentiality
6. Access control
    - Fine-grained Access control
    - Flexible access control

    - Dynamic access control

7. Fairness of information to be shared.

All these factors, including the ownership of the data, copyrights are highly confidential, and also the permissions related factors like access control and user verification has gained importance. The fine-grained has less power, whereas the flexible and dynamic could gain more control over the privacy of data.

Concentrated on the content dissemination factors that are used for analyzing the privacy of data in vehicular social networks where passengers are examined and analyzed using a few factors mentioned below[63]:

Communication behavior
Geographical Location analysis
Individual information analysis
Resistance to counteract the privacy attacks

The communication was based on the posts and information on the social network page, whereas geographical location was based on the latitude and longitude of the social networking devices. If there occur any changes in the individual information as well as location, the attack could be identified. Similarly, the efficiency to hold and reverse the attack from intruders was also recommended in the study.

Few logical factors like robustness, high-fidelity, image sharing scheme from storage were assessed[64] using Facebook as the social networking interface. Image encryption or decryption framework was created to share the images without loss of information and get stored without being assessed by third parties, thereby protecting the data privacy.

Developed a review on various aspects of data privacy, and a different set of factors were identified as given below: Relation with family and friends[65]
Multimedia content in social networks
Genuine users following
Shares from friends
Likes received

The author considered the social networking factors including likes, shares, and relationships with the user as well as users genuinely following a particular user. This would signify the identification of intruder easily as anyone apart from the contacts could be identified anonymous and might be badged as an intruder of privacy information which may be any multimedia content like text, audio, video, animation, or any useful content respectively.

Similar research was conducted[66] where personal factors like friends within the family, education level, hobbies are considered for detecting the privacy of information in social networks. A model was also created, tested, and proved with high performance. Random number generation was applied to protect illegitimate users from entering the system using routing and distributed protocols.

REVIEW FINDINGS

The review performed on various dimensions of Data Privacy Detections in Social Networks (DPDSNs) had different impacts based on social network analysis, data mining, Machine Learning techniques, and Blockchain Technology. Some of the significant findings of the research review are summarised below:

Data Mining was found to be the best technique to design the framework that gives high efficiency in the prediction of data intrusions, violations, and privacy protection mechanisms. Data Mining was also efficient to determine the different stages for prediction of data privacy like pre-processing, cleaning, analysis and detection, etc.

Different algorithms and techniques were used in data mining for efficient prediction of loss of privacy of data in social networks like Ada boost and Naïve Bayesian algorithms.

The reviews on Machine Learning Techniques (MLT) suggested that these techniques could assist in creating the prediction model required for assessing the privacy of information in the best ways possible.

The Machine learning frameworks learned in the reviews was successful in terms of efficiency, reliability as well as the ability to handle online social networking data that was expected to be huge,

Various frameworks were proposed for assessing data privacy in social networks. After the review, it was found that the maximum accuracy of 89.5% was achieved. This showed that data privacy detection should be enhanced with much higher accuracy and performance through the design of the best models.

The frameworks proposed various features for data privacy assessment and were found to be successful. The updated model GOSTDAG also gave good results. Hence it is observed that the blockchain method could be sufficient for social network data privacy detections in online mode. The network access from intruders can be prevented using techniques in blockchain technologies like the Ethereum-based blockchain method or through scenarios like scenario like Ganache, Web3.js, and Remix that was built on the blockchain.

Blockchain techniques were found to be more effective in preserving the privacy of data as well as in protecting the entry from intruders in E-health services.

Different factors related to individual copyrights, family information, personal information, and even location-based factors can be used to assess the privacy of information in social networks using data privatization methods.

The privacy of data was not restricted to text-based information. It also focused on multimedia aspects like image, video, audio, and animation contents that also required privacy in Social networking sites.

After the detailed review analysis of all the techniques, including data mining, MLT, and Blockchain Technology, the impact of each of the methods in acquiring Data Privacy Detection in Social Networks was found commendable.

RESEARCH GAP ANALYSIS

The review of all the different techniques disclosed various factors and methods that can be used to find data privacy violations or address any problems related to it. However, few gaps were identified in the research. The frameworks or any model has not exceeded 90% accuracy in performance in predicting the privacy of data.

The models and frameworks are not sufficient to determine the maximum possible ways to attack the secured privacy data. Lack of Hybrid models which can assess best methods from different sources like the combination of data mining, Machine Learning Technique, and Blockchain Technologies are not available in existence.

Social networking datasets are not available through primary sources in all the reviews conducted, and only pre-existing datasets could be assessed.

Based on the addressed problems, the novel model has to be built with Data privacy Detection algorithms and frameworks possible.

CONCLUSION

The research review studied and presented various techniques and methods applicable to Data Privacy Detection in Social Networks (DPDSNs). The initial thoughts covered the social network researches that covered the data privacy and problems related to it. Later, the data mining followed by machine learning technique-based Data privacy problems in social networks was analyzed. The blockchain technology methods discussed were found to be effective. The summarised analysis of frameworks and factors also quoted the need for a hybrid framework model for future predictions of Data Privacy in Social Networks. This review has steadfastly concluded that the design of the Data Privacy Detection framework should be a hybrid model with the integration of best techniques from three significant technologies viz data mining, machine learning techniques, and blockchain methods respectively.

REFERENCES

[1] Sharma, Sudhir Kumar, Ximi Hoque, and Pravin Chandra. "Sentiment predictions using deep belief networks model for odd-even policy in Delhi." In Cognitive Analytics: Concepts, Methodologies, Tools, and Applications, pp. 1440-1463. IGI Global, 2020.

[2] Derr, Tyler, Zhiwei Wang, JamellDacon, and Jiliang Tang. "Link and interaction polarity predictions in signed networks." Social Network Analysis and Mining 10, no. 1 (2020): 1-14.

[3] Rim, Hyejoon, YoungAh Lee, and SanglimYoo. "Polarized public opinion responding to corporate social advocacy: Social network analysis of boycotters and advocators." Public Relations Review 46, no. 2 (2020): 101869.

[4] Ranjeeth, S., Latchoumi, T. P., Paul, P. V. A Survey on Predictive Models of Learning Analytics. Procedia Computer Science, 2020;167, 37-46.

[5] Colladon, Andrea Fronzetti, and Elisa Remondi. "Using social network analysis to prevent money laundering." Expert Systems with Applications 67 (2017): 49-58.

[6] Ding, Ru-Xi, Xueqing Wang, Kun Shang, and Francisco Herrera. "Social network analysis-based conflict relationship investigation and conflict degree-based consensus reaching process for large scale decision making using sparse representation." Information Fusion 50 (2019): 251-272.

[7] Latchoumi, T. P., Parthiban, L. Abnormality detection using weighed particle swarm optimization and smooth support vector machine.2017

[8] Praveena, A., and S. Smys. "Ensuring data security in cloud based social networks." In 2017 international conference of electronics, communication and aerospace technology (ICECA), vol. 2, pp. 289-295. IEEE, 2017.

[9] Rath, Mamata. "An analytical study of security and challenging issues in social networking as an emerging connected technology." In Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), pp. 26-27. 2018.

[10] Gupta, Brij B., Arun Kumar Sangaiah, Nadia Nedjah, Shingo Yamaguchi, Zhiyong Zhang, and Michael Sheng. "Recent research in computational intelligence paradigms into security and privacy for online social networks (OSNs)." (2018): 851-854.

[11] Latchoumi, T. P., Sunitha, R. Multi agent systems in distributed datawarehousing. In 2010 International Conference on Computer and Communication Technology (ICCCT), 2010; 442-447, IEEE.

[12] Singh, Amit Kumar, Basant Kumar, Sanjay Kumar Singh, S. P. Ghrera, and Anand Mohan. "Multiple watermarking technique for securing online social network contents using back propagation neural network." Future Generation Computer Systems 86 (2018): 926-939.

[13] Loganathan, J., Latchoumi, T. P., Janakiraman, S., parthiban, L. A novel multi-criteria channel decision in co-operative cognitive radio network using E-TOPSIS. In Proceedings of the International Conference on Informatics and Analytics 2016; 1-6.

[14] Pika, Anastasiia, Moe T. Wynn, Stephanus Budiono, Arthur HM ter Hofstede, Wil MP van der Aalst, and Hajo A. Reijers. "Towards privacy-preserving process mining in healthcare." In International Conference on Business Process Management, pp. 483-495. Springer, Cham, 2019.

[15] Li, Guang. "A new bayesian-based method for privacy-preserving data mining." In International Conference on Intelligent and Interactive Systems and Applications, pp. 171-177. Springer, Cham, 2017.

[16] Babu, Pasupuleti Nagendra, and S. Ramakrishna. "Critical Review on Privacy and Security Issues in Data Mining." In Emerging Research in Data Engineering Systems and Computer

Communications, pp. 217-230. Springer, Singapore, 2020.

[17] KUMAR, D., and GOPU NAGARJUNA REDDY. "Controlling Health Admissions By Using Data Mining Methods."

[18] Parkhimenka, Uladzimir, Mikhail Tatur, and Anna Zhvakina. "Heuristic approach to online purchase prediction based on internet store visitors classification using data mining methods." In 2017 International Conference on Information and Digital Technologies (IDT), pp. 304-307. IEEE, 2017.

[19] Latchoumi, T. P., Ezhilarasi, T. P., Balamurugan, K. Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 2019;1(10), 1137.

[20] Askinadze, Alexander, and Stefan Conrad. "Respecting Data Privacy in Educational Data Mining: An Approach to the Transparent Handling of Student Data and Dealing with the Resulting Missing Value Problem." In 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), pp. 160-164. IEEE, 2018.

[21] Noaman, Amin Y., Farrukh Nadeem, Abdul Hamid M. Ragab, Arwa Jamjoom, Nabeela Al-Abdullah, Mahreen Nasir, and Anser G. Ali. "Improving Prediction Accuracy of “Central Line-Associated Blood Stream Infections” Using Data Mining Models." BioMed research international 2017 (2017).

[22] Schutte, Jeanine, Alta Van Der Merwe, and FransonetReyneke. "Using Data Analytics and Data Mining Methods to Determine a High Net Worth IndividualÃ¢ Â Â s Electronic Banking Behavior." Journal of Internet Banking and Commerce 22, no. 3 (2017): 1-39.

[23] Scardapane, Simone, Rosa Altilio, Valentina Ciccarelli, Aurelio Uncini, and Massimo Panella. "Privacy-preserving data mining for distributed medical scenarios." In Multidisciplinary Approaches to Neural Computing, pp. 119-128. Springer, Cham, 2018.

[24] Kretschmer, Ralf, Alain Pfouga, Stefan Rulhoff, and J. Stjepandić. "Knowledge-based design for assembly in agile manufacturing by using Data Mining methods." Advanced Engineering Informatics 33 (2017): 285-299.

[25] Sohrabi, Mohammad Karim, and Soodeh Akbari. "A comprehensive study on the effects of using data mining techniques to predict tie strength." Computers in Human behavior 60 (2016): 534-541.

[26] Kotenko, Igor, Igor Saenko, and Andrey Chechulin. "Protection against information in eSociety: using Data Mining methods to counteract unwanted and malicious data." In International Conference on Digital Transformation and Global Society, pp. 170-184. Springer, Cham, 2017.

[27] Chamikara, MahawagaArachchige Pathum, Peter Bertók, Dongxi Liu, SeyitCamtepe, and Ibrahim Khalil. "Efficient privacy preservation of big data for accurate data mining." Information Sciences 527 (2020): 420-443.

[28] Gaur, Mithun. "Privacy Preserving Machine Learning Challenges and Solution Approach for Training Data in ERP Systems." International Journal of Computer Engineering and Technology (2020).

[29] Li, Qinbin, Zeyi Wen, and Bingsheng He. "Federated learning systems: Vision, hype and reality for data privacy and protection." arXiv preprint arXiv:1907.09693 (2019).

[30] Latchoumi, T. P., Balamurugan, K., Dinesh, K., & Ezhilarasi, T. P. (2019). Particle swarm optimization approach for waterjet cavitation peening. Measurement, 141, 184- 189.

[31] Rodríguez-Barroso, Nuria, Goran Stipcich, Daniel Jiménez-López, José Antonio Ruiz- Millán, Eugenio Martínez-Cámara, Gerardo González-Seco, M. Victoria Luzón, Miguel Angel Veganzones, and Francisco Herrera. "Federated Learning and Differential Privacy: Software

tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy." Information Fusion 64 (2020): 270-292.

[32] Nasr, Milad, Reza Shokri, and Amir Houmansadr. "Machine learning with membership privacy using adversarial regularization." In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, pp. 634-646. 2018.

[33] Tesfay, Welderufael B., Peter Hofmann, Toru Nakamura, ShinsakuKiyomoto, and Jetzabel Serna. "I read but don't agree: Privacy policy benchmarking using machine learning and the eugdpr." In Companion Proceedings of The Web Conference 2018, pp. 163-166. 2018.

[34] Xiao, Liang, Xiaoyue Wan, Xiaozhen Lu, Yanyong Zhang, and Di Wu. "IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?" IEEE Signal Processing Magazine 35, no. 5 (2018): 41-49.

[35] Ranjeeth, S., Latchoumi, T. P., & Victer Paul, P. (2019). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications. https://doi. org/10.2174/2666255813666191116150319.

[36] Asok, Divya, P. Chitra, and BharathirajaMuthurajan. "Privacy Preserving Machine Learning and Deep Learning Techniques: Application–E-Healthcare." In Handbook of Research on Applications and Implementations of Machine Learning Techniques, pp. 222-235. IGI Global, 2020.

[37] Tanuwidjaja, Harry Chandra, Rakyong Choi, SeunggeunBaek, and Kwangjo Kim. "Privacy-Preserving Deep Learning on Machine Learning as a Service—a Comprehensive Survey." IEEE Access 8 (2020): 167425-167447.

[38] Dagher, Gaby G., Jordan Mohler, MateaMilojkovic, and Praneeth Babu Marella. "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology." Sustainable cities and society 39 (2018): 283-297.

[39] Biswas, Kamanashis, and VallipuramMuthukkumarasamy. "Securing smart cities using blockchain technology." In 2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS), pp. 1392-1393. IEEE, 2016.

[40] da Conceição, Arlindo Flavio, Flavio Soares Correa da Silva, Vladimir Rocha, Angela Locoro, and João Marcos Barguil. "Eletronic health records using blockchain technology." arXiv preprint arXiv:1804.10078 (2018).

[41] Rifi, Nabil, Elie Rachkidi, Nazim Agoulmine, and Nada Chendeb Taher. "Towards using blockchain technology for eHealth data access management." In 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), pp. 1-4. IEEE, 2017.

[42] Wang, Hao, Shenglan Ma, Hong-Ning Dai, Muhammad Imran, and Tongsen Wang. "Blockchain-based data privacy management with nudge theory in open banking." Future Generation Computer Systems 110 (2020): 812-823.

[43] Loukil, Faiza, ChirineGhedira-Guegan, KhouloudBoukadi, and Aïcha Nabila Benharkat. "Towards an end-to-end IoT data privacy-preserving framework using blockchain technology." In International Conference on Web Information Systems Engineering, pp. 68-78. Springer, Cham, 2018.

[44] Dang, Thanh Long Nhat, and Minh Son Nguyen. "An approach to data privacy in smart home using blockchain technology." In 2018 International Conference on Advanced Computing and Applications (ACOMP), pp. 58-64. IEEE, 2018.

[45] Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). Role of gender on academic performance based on different parameters: Data from secondary school education. Data in brief, 29, 105257.

[46] Chakraborty, Sabyasachi, SatyabrataAich, and Hee-Cheol Kim. "A secure healthcare system design framework using blockchain technology." In 2019 21st International Conference on Advanced Communication Technology (ICACT), pp. 260-264. IEEE, 2019.

[47] Rantos, Konstantinos, George Drosatos, Konstantinos Demertzis, Christos Ilioudis, Alexandros Papanikolaou, and AntoniosKritsas. "ADvoCATE: a consent management platform for personal data processing in the IoT using blockchain technology." In International Conference on Security for Information Technology and Communications, pp. 300-313. Springer, Cham, 2018.

[48] Joshi, Archana Prashanth, Meng Han, and Yan Wang. "A survey on security and privacy issues of blockchain technology." Mathematical foundations of computing 1, no. 2 (2018): 121.

[49] Srivastava, Gautam, Ashutosh Dhar Dwivedi, and Rajani Singh. "Automated remote patient monitoring: data sharing and privacy using blockchain." arXiv preprint arXiv:1811.03417 (2018).

[50] Arora, Dev, Siddharth Gautham, Harshit Gupta, and Bharat Bhushan. "Blockchain- based Security Solutions to Preserve Data Privacy and Integrity." In 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 468-472. IEEE, 2019.

[51] Wang, Haoxiang. "IoT based Clinical Sensor Data Management and Transfer using Blockchain Technology." Journal of ISMAC 2, no. 03 (2020): 154-159.

[52] Zhang, Mian, and Yuhong Ji. "Blockchain for healthcare records: A data perspective." PeerJ Preprints 6 (2018): e26942v1.

[53] Zheng, Xu, Zhipeng Cai, Guangchun Luo, Ling Tian, and Xiao Bai. "Privacy-preserved community discovery in online social networks." Future Generation Computer Systems 93 (2019): 1002-1009.

[54] Ahamed, B. Bazeer, and D. Yuvaraj. "Framework for faction of data in social network using link based mining process." In International Conference on Intelligent Computing & Optimization, pp. 300-309. Springer, Cham, 2018.

[55] Valliyammai, Chinnaiah, and AnbalaganBhuvaneswari. "Semantics-based sensitive topic diffusion detection framework towards privacy aware online social networks." Cluster Computing 22, no. 1 (2019): 407-422.

[56] Song, Xuemeng, Xiang Wang, LiqiangNie, Xiangnan He, Zhumin Chen, and Wei Liu. "A personal privacy preserving framework: I let you know who can see what." In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 295-304. 2018.

[57] Samarah, Samer, Mohammed Gh Al Zamil, Ahmed F. Aleroud, MajdiRawashdeh, Mohammed F. Alhamid, and Atif Alamri. "An efficient activity recognition framework: Toward privacy-sensitive health data sensing." IEEE Access 5 (2017): 3848-3859.

[58] Sonone, Karuna, and V. M. Barkade. "APRIGUARD: The Automatic and Eficient Method for Privacy Violations Detection in OSNs (Online Social Networks." In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), pp. 945-949. IEEE, 2018.

[59] Soliman, Amira, and Sarunas Girdzijauskas. "Dlsas: Distributed large-scale anti-spam framework for decentralized online social networks." In 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC), pp. 363-372. IEEE, 2016.

[60] Muhammad, Khan, Muhammad Sajjad, Irfan Mehmood, Seungmin Rho, and Sung WookBaik. "Image steganography using uncorrelated color space and its application for security of visual contents in online social networks." Future Generation Computer Systems 86 (2018): 951-960.

[61] Audeh, Bissan, Michel Beigbeder, Antoine Zimmermann, Philippe Jaillon, and Cédric Bousquet. "Vigi4Med scraper: A framework for web forum structured data extraction and semantic representation." PloS one 12, no. 1 (2017): e0169658.

[62] Malekhosseini, Razieh, Mehdi Hosseinzadeh, and Keyvan Navi. "An investigation into the requirements of privacy in social networks and factors contributing to users' concerns about violation of their privacy." Social Network Analysis and Mining 8, no. 1 (2018): 41.

[63] Wang, Xiaojie, Zhaolong Ning, MengChu Zhou, Xiping Hu, Lei Wang, Yan Zhang, Fei Richard Yu, and Bin Hu. "Privacy-preserving content dissemination for vehicular social networks: Challenges and solutions." IEEE Communications Surveys & Tutorials 21, no. 2 (2018): 1314-1345.

[64] Sun, Weiwei, Jiantao Zhou, Shuyuan Zhu, and Yuan Yan Tang. "Robust privacy- preserving image sharing over online social networks (osns)." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14, no. 1 (2018): 1-22.

[65] Rathore, Shailendra, Pradip Kumar Sharma, Vincenzo Loia, Young-SikJeong, and Jong Hyuk Park. "Social network security: Issues, challenges, threats, and solutions." Information sciences 421 (2017): 43-69.

[66] Chen, Yu, Hanchao Ku, and Mingwu Zhang. "PP-OCQ: A Distributed Privacy- Preserving Optimal Closeness Query Scheme for Social Networks." Computer Standards & Interfaces (2020): 103484.