A Privacy Preserving Framework to Protect Sensitive Data in Online Social Networks

Nisha P. Shetty, Balachandra Muniyal*, Niraj Yagnik, Tulika Banerjee and Angad Singh

Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal-567104, India
E-mail: bala.chandra@manipal.edu
**Corresponding Author*

Received 24 November 2021; Accepted 05 May 2022;
Publication 07 November 2022

Abstract

In this day and age, Internet has become an innate part of our existence. This virtual platform brings people together, facilitating information exchange, sharing photos, posts, etc. As interaction happens without any physical presence in the medium, trust is often compromised in all these platforms operating via the Internet. Although many of these sites provide their ingrained privacy settings, they are limited and do not cater to all users' needs. The proposed work highlights the privacy risk associated with various personally identifiable information posted in online social networks (OSN). The work is three-facet, i.e. it first identifies the type of private information which is unwittingly revealed in social media tweets. To prevent unauthorized users from accessing private data, an anonymous mechanism is put forth that securely encodes the data. The information loss incurred due to anonymization is analyzed to check how much of privacy-utility trade-off is attained. The private data is then outsourced to a more secure server that only authorized people can access. Finally, to provide effective retrieval at the server-side, the traditional searchable encryption technique is modified,

Journal of Cyber Security and Mobility, Vol. 11_4, 575–600.
doi: 10.13052/jcsm2245-1439.1144

considering the typo errors observed in user searching behaviours. With all its constituents mentioned above, the purported approach aims to give more fine-grained control to the user to decide who can access their data and is the correct progression towards amputating privacy violation.

Keywords: Data anonymization, social media privacy, secure searchable encryption, personally identifiable information.

1 Introduction

The dawn of various social media channels has revolutionized the internet landscape by making global networking reach every home. Today, connecting to people worldwide living in different time zones is a matter of a few seconds. Online Social Networks (OSN) has become a mass phenomenon in the early 20th century due to their new cost-effective radical way of sharing interests and activities. Mark Zuckerberg, the creator of Facebook, was once quoted saying that the appeal of social networks comes from the fact that it offers a way to "stay connected" with others.

A classic OSN offers its users a simulated environment, governed by their policies, to share their data, make and interact with friends. Users can also share data in their friends' space or tag another user providing a link to his personal virtual space in their profile. The two stakeholders in Online Social Network privacy are; the users who share information and OSN, which manages users' accounts online and is responsible for providing good continuous services. OSN functionality can be categorized in the following fashion [1]:

- The networking functions enable users to cultivate relationships in the virtual scenario.
- User-provided content and interactions fall under data functions. Some of these contents include personally identifiable information such as birthday, email, phone number, bio, marital status, etc. [2].
- Access control functions regulate and administer user-defined privacy settings and rules.

However, with all these advantages, there are rising concerns observed in the privacy and protection of personal data. Threat to privacy can be categorized in the following manner [3, 4]:

- Privacy Breach: Gathering crucial information about a person such as personally identifiable information, connections between people, etc. and exploiting them.

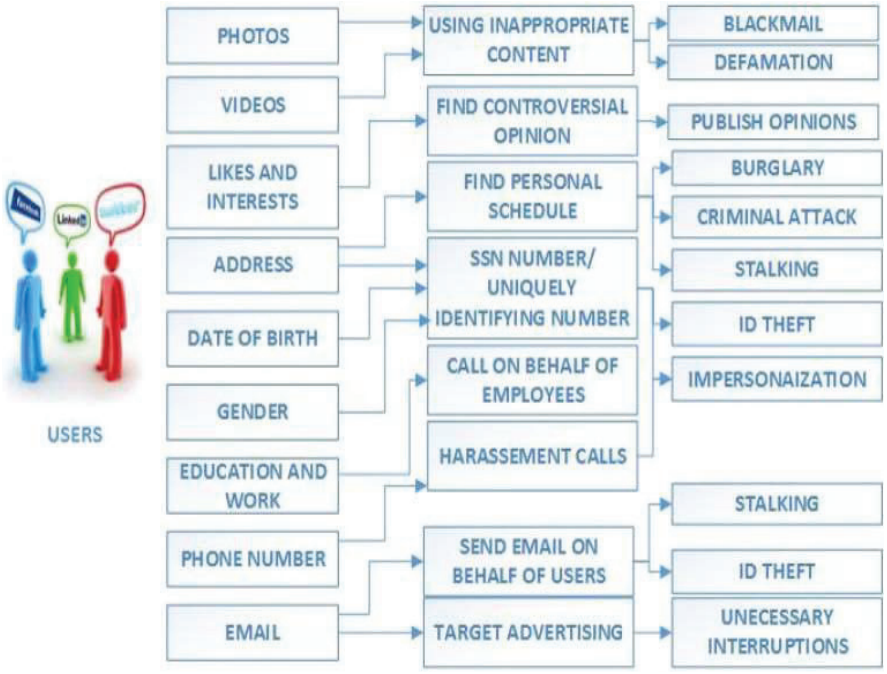


Figure 1 Effects of disclosure of sensitive information on data privacy [5].

- Passive Attack: Undetectable attacks are primarily made for research and advertising purposes.
- Active Attacks: Exploiting friendship links of users and other potentially harmful information to cause harm.

Some of the effects of such disclosures are listed in Figure 1.

All such activities have elevated the level of risk and have made users vulnerable to any malicious manipulation [6, 7]. Often the existing OSN regulatory mechanisms fail to cater to users' requirements, and so studies on determining trust while anonymizing sensitive data in OSN have gained the interest of many researchers worldwide.

Anonymization of data involves the techniques of making the data incognito so that private information can no longer be identified [8]. It is defined as encrypting or removing personally identifiable information detected in the dataset, thus concealing the people's identity. This technique eliminates any attempt to re-identify the personal data that has been made indiscernible to the end user's eyes.

Anonymization techniques can be categorized broadly as [9]:

- Generalization: Deals with replacing the actual value with a higher-order magnitude value.
- Randomization: Obscuring the actual values by the addition of noise.

Pseudonymization is different from anonymization as the sensitive parts of the data here are replaced so that the private information can be re-identified using an appropriate additional identifier.

In the proposed research, the following operations are incorporated: masking with hash values, masking with fake data and obfuscating the attributes via differential privacy. The contributions offered by the proposed work can be summarized as follows. A framework that employs machine learning classifiers to detect PII or any sensitive data revealed in the tweets is designed. A computationally efficient module is hypothesized to encipher such crucial data ensuring semantic coherence of the sentence.

The following sections detail recent studies in the related domain in Section 2; proposed methodology in Section 3; result analysis, discussions, and conclusion in Sections 4, 5 and 6.

2 Related Works

2.1 Data Anonymization

A growing body of literature is available in the domain of data anonymization. A few of the key most relevant works are listed below.

An increasing number of works has been conducted in anonymizing geospatial data of users. In one such notable study, authors, Hasanzadeh et al. [10] (2020) presented an anonymization technique comprising k-anonymity and Gaussian displacement algorithm. The main limitation of this work is a failure against background knowledge attack even if the records are made 'k' indiscernible. Stricter anonymization policies must be put forth to achieve the privacy of sensitive information such as medical data.

Lisin and Zapechnikov (2020) [11] discussed two main approaches to privacy-preserving machine learning (cryptographic and perturbation), together with examples of how to use some of these methods in practice.

Gaur (2020) [12] focused on the critical challenges faced by ERP companies while training machine learning models on private enterprise data. The work examines the role of anonymization and differential privacy in protecting sensitive data. The work does not, however, consider sensitive political and religious data. In our work, we adapted the algorithms on the basis of the author's observations.

Siddula et al. (2019) [13] introduced a clustering-based anonymization method which ensures identity, contacts and attribute privacy by implementing a k-anonymity on the users sharing similar attributes. The proposed work ensures sensitive attribute privacy through improved l-diversity. Although innovative, the method can be however be enhanced by incorporating ways to reduce inter-cluster distance. This can be achieved by filtering an outright number of users in a cluster based on the right attribute combination.

A fundamental problem in a generalization-based method is replacing the actual value by over fitted/under fitted intervals. Abdul Majeed (2019) [14] outlined a classical approach wherein users are grouped based on their similarity, ranking them into appropriate equivalence classes with proper attribute range analysis. The author substituted numerical and categorical values with their right counterpart mean and IDs. However, the important limitation observed in this study was that it was limited to only a few sensitive attributes. Taking into account multiple diverse characteristics and their interrelationships can be a notable research direction in future.

Experiments to improve two-party secure computing protocol using a hybrid technique involving association rule mining and homomorphic encryption were put forth by authors Ouyang and Huang (2019) [15]. Unfortunately, appropriate qualitative analysis is not performed on encrypted data to analyze its utility. The lack of optimization protocols in such a computationally intensive multiparty scenario is a major flaw of their experiment.

M. Dias, A. Abad and I. Trancoso (2018) [16] focused on creating privacy-preserving techniques in the context of a speech emotion recognition task as a proof of concept that might be applied to other speech analytics projects. The proposed work used homomorphic encryption and distance-preserving hashing techniques to successfully protect sensitive data with minimal degradation costs in terms of predictive model accuracy. Although very effective, the proposed method is computationally demanding. Authors recommend the implementation of intricate classifiers based on differential privacy to further their research.

Authors Wei et al. (2018) [17] investigated the pitfalls of various data protection methods in social media. They put forth an innovative amalgamation. The HHGA-RBF neural network algorithm examined the security state in OSN, followed by SVM pre-processing the data, ABES encrypting, and finally, PSO improving the security circumstances. The highly comprehensive method put forth by the authors proved computationally more efficient than classic encryption techniques. The technique proved to be more robust than traditional techniques in terms of information loss and privacy

protection. The authors have outlined the possible future ventures, such as incorporating graphical and chaotic encryption to further their study.

In their research, authors Macwan and Patel [18] (2018) tried to counteract the Mutual friend's attack by proposing an improvement to the k-anonymity method. Their graph modification technique added more edges to modify the structural information between the nodes while ensuring that the number of vertices remained the same as the original data set. To further their research, the authors plan to propose the solution to the neighbourhood attack too.

In their analysis, the author's Yuan et al. [19] (2013) claimed that most of the existing graph modification techniques distort the graph properties by insertion/deletion of edges. To address this issue, the authors proposed a method to introduce noise nodes to provide anonymity. There is still a possibility of achieving a better trade-off between the number of introduced noise nodes and the amount of anonymization reached.

2.2 Differential Privacy

More work on the potential of differential privacy as a critical component in user identity protection has been carried out lately. However, its scope in obfuscating sensitive node attributes is vastly left unexplored.

Huang et al. (2020) [20] traced the advances of application of differential privacy for preserving privacy in their work. They proposed a combination of differential privacy with randomness and clustering to balance data availability and the right level of protection. The complexity of the approach, however, poses a considerable challenge in its applicability in large networks.

N. Wu et al. [21] (2019) applied machine learning to private data owned by multiple distributed owners. The authors achieved an excellent optimization in their technique and focused our attention on the difference observed in the fitness of a model when trained with differentially private queries. The authors recommend further studies in the same domain targeting adversarial learning scenarios.

For the apt application of Differential Privacy Library, Holohan et al. (2019) in [22] aided us in correctly understanding the foundations of differential privacy and its various application scenarios. This work's main contribution is to provide a unified code base that can be used for future works in the domain.

Experiments conducted by Triastcyn and Faltings (2019) [23] put forth a cutting-edge method that was able to train faster, not susceptible to outliers

and achieved good privacy with very little noise addition. Their proposed Bayesian approach was, however, proven effective only for finite datasets.

Xu et al. [24] (2019) implemented fairness aware fusion of differential privacy and logistic regression to achieve similarity amongst protected and unprotected groups. To some extent, our approach was inspired by this combination. The proposed method was able to acquire good data utility. However, based on the sensitivity of the attribute, different noise coefficients can be added to achieve varying levels of privacy.

2.3 Secure Searchable Encryption

Chen et al. (2020) [25] incorporated the search mechanism with a memorable user password. Although this method was proven computationally efficient concerning key management, it increased the design overhead while creating a distributed architecture to prevent dictionary attacks.

Ahsan et al. (2018) [26] introduced a vigilant scheme against keyword guessing attacks. In their method, ciphertexts of the search keywords are sent along with encrypted emails to the server. On receiving the cipher keywords, the receiver decrypts them and responds with the REST keywords. The server then stores the REST keywords validated against the trapdoor of the keyword sent by the receiver to grant access to the relevant emails. Thus, their method ensures authentication without actually revealing the keyword. Further, the authors aim to incorporate multiple keywords with fuzzy search and introduce a ranking mechanism with a priority feature to sort the search results relevantly.

3 Methodology

A framework to extract and anonymize crucial information while effectively reducing the computational overhead observed in base symmetric searchable encryption due to human errors is put forth in the following sections, as illustrated in Figure 2.

3.1 Data Generation

3.1.1 Extraction of sensitive religious and political data

The proposed work uses 'Tweepy API' [27] to extract the latest tweets to generate a religious and political data-sensitive corpus. Tweets are scraped from the site to construct the needed database using relevant hashtags pertaining to

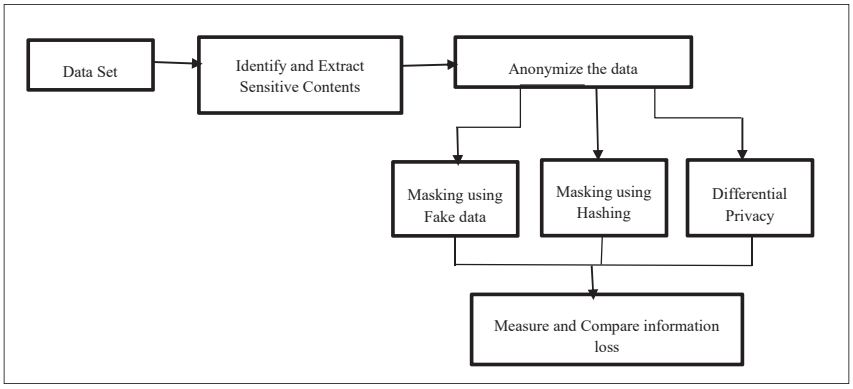


Figure 2 Overall methodology.

religion and politics. Two thousand five hundred tweets are extracted for both politics and religion. This database is labelled as sensitive data.

3.1.2 Generation of private data

Since it is tough to get hold of any openly available private database, the proposed work uses faker [28] to generate fake sensitive texts. A comprehensive database with text data is developed having personal and confidential information like Name, Address, Social Security Number, Email, Credit Card Number, Phone Number, Date of Birth and Home Address [29]. A database of 9000 rows is generated containing personal, private data. This data is labelled as private data.

3.1.3 Data fusion

The above generated two databases are combined with 1500 data entries that contain text data, which are neither private nor sensitive. These entries are labelled as None. Finally, we have a database with text data with three types of data (as shown in Figure 3). The dataset includes text data embedded with sensitive religious/political data, text data with personally identifiable information, and text data devoid of any sensitive or confidential information.

3.1.4 Sentiment association

The proposed work uses Valence Aware Dictionary for Sentiment Reasoning, or Vader [30] for associating a positive or negative sentiment to the text database created. Sentiments give the actual impression of the person on a particular topic.

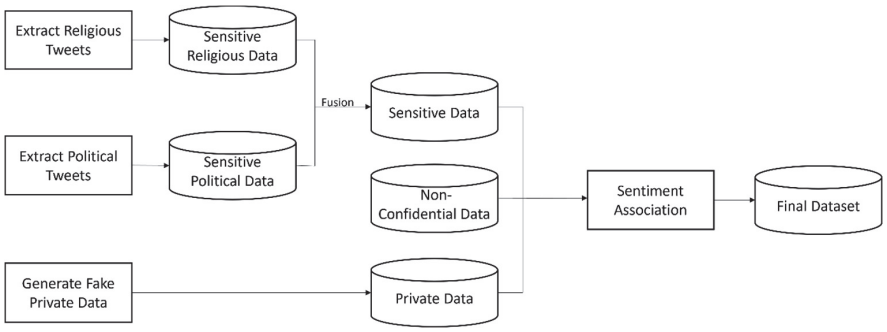


Figure 3 Data set generation.

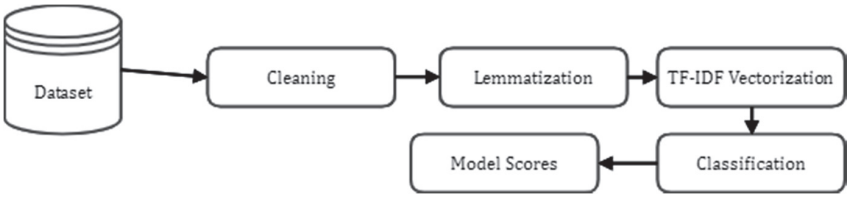


Figure 4 Pipeline to identify and extract sensitive contents.

3.2 Identify and Extract Sensitive and Private Contents

3.2.1 Identification of data type

This segment emulates the process that an organization can use to identify which data in the database are private or sensitive to be processed accordingly before being outsourced. The proposed work exercises machine learning models to classify the data based on training them using the text dataset of the target variable's data type, as shown in Figure 4.

3.2.2 Pre-processing

Before implementing the classification algorithms, pre-processing is carried out on the prepared dataset. Any HTML tags, English stop words or non-alphabetic characters that may be in the dataset are removed. For uniformity, all of the text is also converted to lowercase. Finally, the proposed work applies the Term Frequency-Inverse document frequency (TF-IDF) vectorizer followed by lemmatization to the data. The vectorizer lets us obtain a set of features, whereas lemmatization helps decrease the number of redundant features in the set.

3.2.3 Methodology and implementation of classification of data type

The dataset is divided into training and test sets. We implement five classification models - Random Forest, Support Vector Classifier, Logistic Regression, Decision Tree and XGB Classifier to classify the data as either private information, sensitive information, or neither.

3.2.4 Extraction of private/sensitive data

A module is developed which, when presented with a text categorized as Sensitive/Private text [31], can identify the exact private and personal information embedded in the text. The module uses regex functions and named entity recognition techniques to extract the relevant categories of confidential data. This extracted information will be essential for the future modules of the methodology for masking this information for the data mining tasks.

3.3 Anonymization/Pseudonymization Techniques

3.3.1 Differential privacy

An algorithm is differentially private [22]; if the person's PII contributing to the output cannot be inferred. Differential privacy is proven to be efficient against background knowledge attacks. This has proven most effective against probing researchers by facilitating them to unearth the patterns in OSN users' behaviour while obscuring the information about each individual's records.

An algorithm A is ϵ-differentially private if and only if:

$$\Pr[\mathrm{A}(\mathrm{D}) = \mathrm{x}] \leq \mathrm{e}^{\wedge}(\epsilon) * \Pr[\mathrm{A}(\mathrm{D}') = \mathrm{x}] \tag{1}$$

Extrapolating this concept, our model (as shown in Figure 5) aims to take the private data that individuals may unknowingly put on their social media feeds, anonymize it and replace it with certain randomized fake data (by adding noise) while maintaining coherence, and then send it back, thereby preserving any private information about the individual that could potentially be misused.

The proposed work uses the Diffprivlib [22] library by IBM to provide an easy and efficient medium to explore the impact of differential privacy on machine learning accuracy while performing classification. The effect of differential privacy is studied on the Logistic Regression model after the textual data is cleaned, lemmatized, and vectorized, making it fit for training.

Figure 5 Differential privacy pipeline.

3.3.2 Anonymization using fake data

The proposed work anonymises sensitive textual information by masking it with fake system-generated texts to protect the user's confidential information from external threats. The proposed work uses artificial data generation for hiding the original private data to protect them. Faker API [28] generates the simulated data corresponding to the PII while maintaining the sentence's logical coherence.

3.3.3 Pseudonymization using hashing

A cryptographic hash function is a mathematical process that returns a fixed-size bit string from an arbitrary data block. This technique can be used for various security-related tasks, including file comparison, blockchain verification and hiding sensitive data. The proposed work uses SHA256 as the cryptographic hashing function to mask the identified private data in the text, thus making reverse engineering or detection impossible. Since the hash value is unique for each text-only authorized users with correct keys can decipher the text.

3.4 Searchable Symmetric Encryption

Encryption or hashing often converts the data into a random format which makes the search function infeasible. Usually, such texts must be decrypted overall to aid the search function making it computationally expensive. Searchable encryption enables searching in the 'ciphertext format' with minimum data leakage.

The masking of the private data is topped off with an architecture of Searchable Symmetric Encryption (SSE), allowing a party to outsource the data to another party in a confidential manner while still maintaining the ability for the party to search for an entry selectively. The data is encrypted locally by the client party and is exported to the service provider (SP), with the SP having no information about the encryption key.

Each of the crucial data pertaining to each user is stored in encrypted files and associated with a set of keywords. To fetch the data file, the authorized user needs to type the keyword. The server then returns the set of files bracketed under the inputted keyword. However, an added advantage of this approach [32] is that in the case of typos in the search query, the server employs predefined semantic metrics and returns the closest results. The augmented work uses edit distance to quantify keywords similarity to provide a secure and privacy-preserving solution.

Edit distance preliminary: The no. of the operations (substitution, deletion or insertion of a character) needed to transform one word into another is called edit distance.

Mathematically, the whole process can be represented in the following fashion:

For n encrypted files in server F-W = (F_a–w_1,... F_z–w_N) where F is the file name linked to specific keyword w with a pre-determined edit distance d, if the search input is (w_i, k) the server analyses in the following manner:

if wi ∈ F-W then
the server returns the IDS of all F_i associated with w_i
else
the server returns the IDS of all F_z wherein ed(w_i, w_z) $\leq$ k

3.5 Post Masking Analysis

Figure 6 illustrates the post masking scenario.

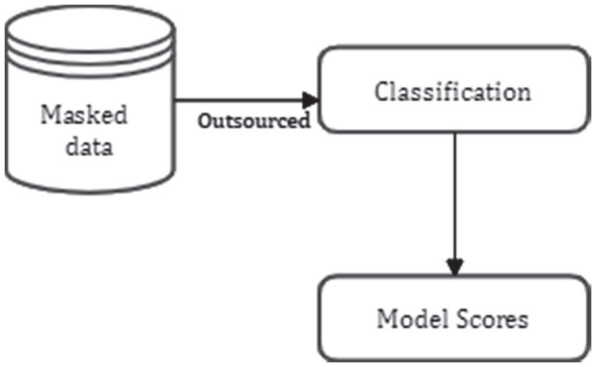


Figure 6 Evaluation post anonymization.

4 Result Analysis and Discussion

The proposed methodology is tested on the created dataset consisting of private and sensitive data and associated sentiment to check its efficacy and consistency. After pre-processing and vectorizing the textual contents, the results obtained from the classification algorithms are recorded to compare model performances and the applicability of the proposed methodology. XGBoost, Logistic Regression, Random Forest, Decision Tree and Support Vector Classifier are used for training the models. The models are trained on three versions of the database. The models are initially trained to identify unprotected text data. The models are then trained on the hashed and pseudonymized versions of the text database. Any differences in the model performances are studied to discuss the robustness of the method proposed.

4.1 Observations and Illustrations

1. Data Type identification pre and post masking – First, the supervised models are trained to predict the type of private data the particular textual sentence entails. The high accuracies obtained by the models indicate the high efficacy of the data-type identifier and their reliability while trying to predict the type of private information depicted in the text. Models are then trained on the secure database with hashed and pseudonymized data. The results obtained (Figure 7, Tables 1 and 2) show that despite hiding the private information, the model accuracies can be maintained if the logical coherence of the textual data can be maintained.

2. Differential Privacy – The effect of differential privacy on the Logistic Regression model performance is studied in the proposed work. Differential privacy is first performed on the Logistic Regression model using an epsilon value of 1.0, indicating a high degree of noise introduced to the private data. Further experiments are performed for a range of epsilon values to study the effect of epsilon on the model accuracies and security offered. The model score (accuracy) depreciates considerably when differential privacy is incorporated with a low epsilon value. As the epsilon is increased, the model scores start to increase as the noise introduced decreases. Thus, indicating the need of reaching the right and required compromise in the accuracy-privacy trade-off while training the models, as shown in Figures 8 and 9.

3. Modified Secure Searchable Encryption (SSE): Below a sample output for SSE is illustrated.

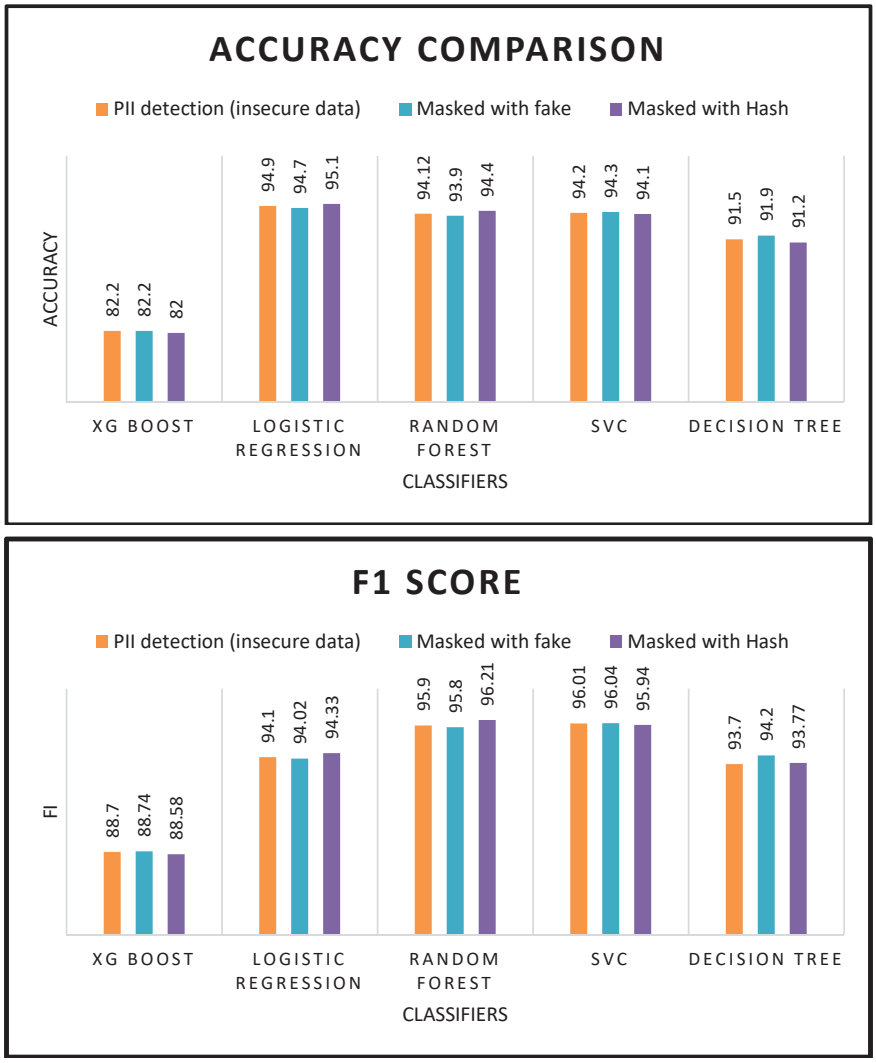


Figure 7 Continued

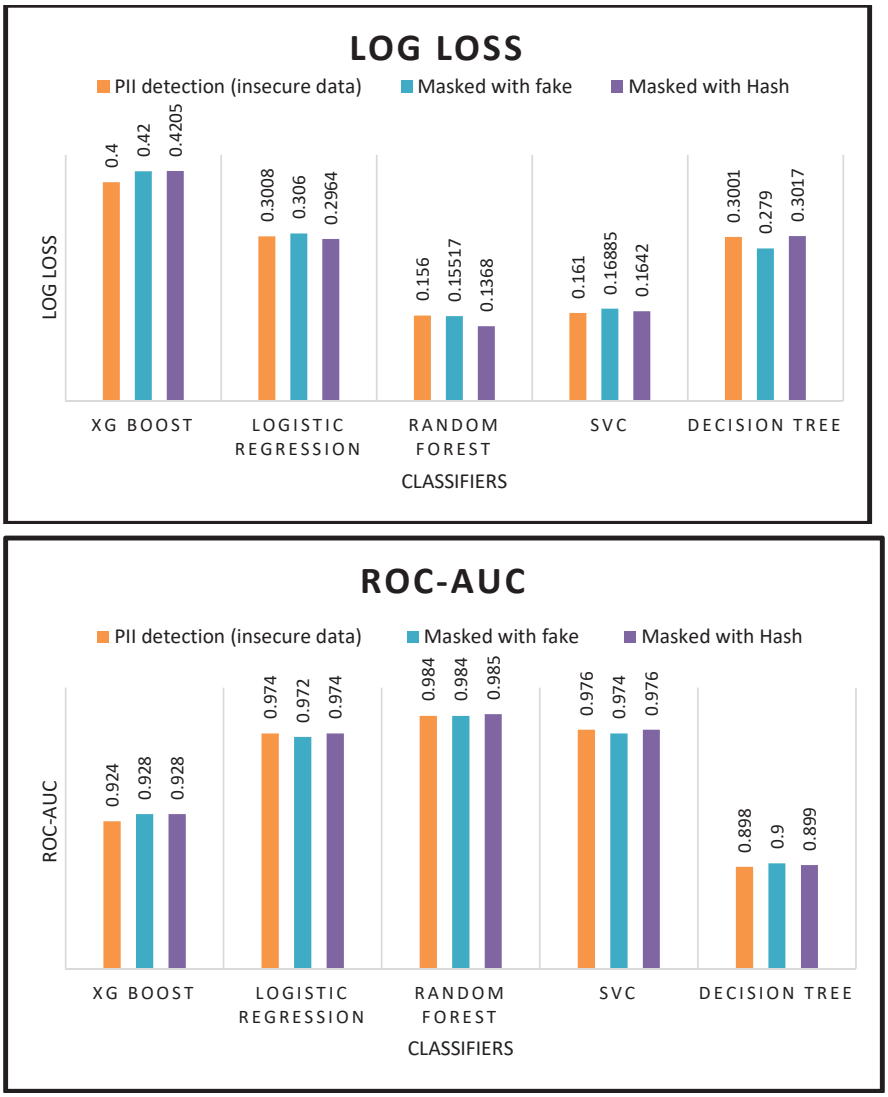


Figure 7 Performance Metrics of Classifiers on Insecure, Fake and Hashed data.

Table 1 Performance evaluation – Cohen kappa

	XGBoost	Random Forest	Support Vector Classifier	Decision Tree	Logistic Regression
Masked with Hash	0.76	0.85	0.778	0.79	0.784
Masked with Fake data	0.487	0.851	0.782	0.8041	0.481

Table 2 Performance evaluation – Mathews correlation coefficient

	XGBoost	Random Forest	Support Vector Classifier	Decision Tree	Logistic Regression
Masked with Hash	0.781	0.860	0.765	0.791	0.795
Masked with Fake data	0.553	0.8561	0.7794	0.8041	0.547

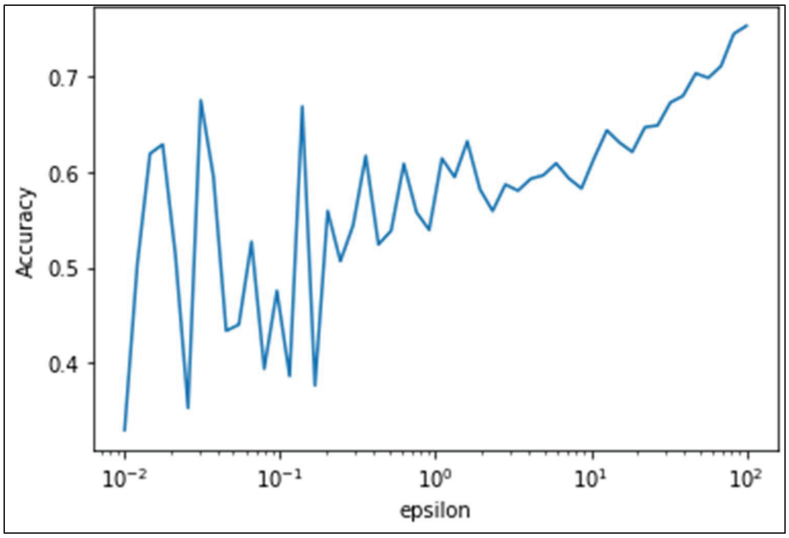


Figure 8 Effects of Epsilon value on accuracy.

Enter your query: Nepam
A set of fuzzy keywords:
*['*Nepam', '*epam', 'N*pam', 'Ne*am', 'Nep*m', 'Nepa*', 'Nepam*']*
Searching for results...
Server returned these encrypted file identifiers:
[b'gAAAAABg8Ra_wycYudzStA9gk28r7Q4MBOsPjxL2GFRXDiGU_c7281P
rqRWvfq4f5aT2OB3Ty_kOP78fRhOVggWZoLKzKzcT5w==',
b'gAAAAABg8RbAmMbAO53mNqLa27IUJLKx8SItUd998dnZYahzAOOGRi
MLWAqQqWwS5rtLDpSpZ3uiPvdnFymGXh68cKVnaAlgPA==',

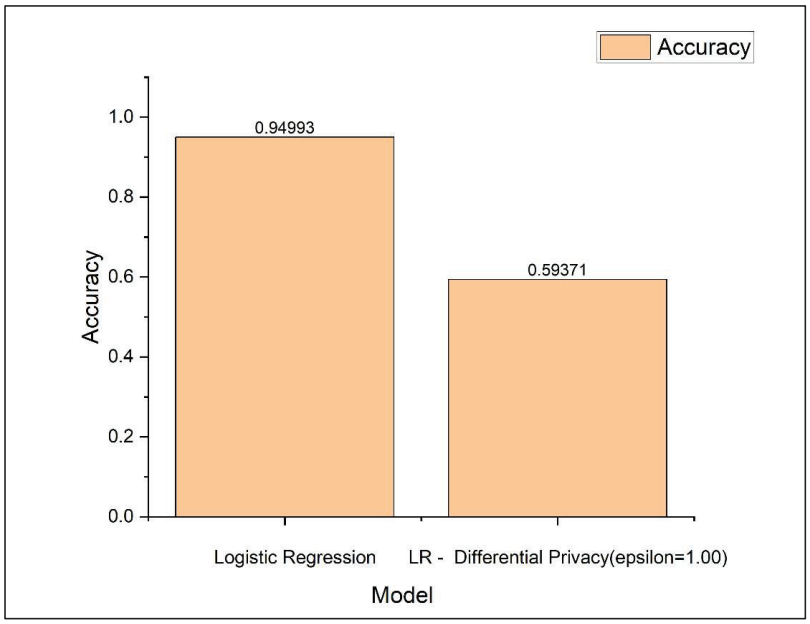


Figure 9 Model accuracy scores with and without differential privacy.

b'gAAAAABg8RbAB40Dw5_1CDXqg5b_tVZYvXkr6KW93J7VSoN2sxb0wxb OVoO4zdkorZo4MZ-vVwKQD3cMUxS7Yq3eOM0wNsdytA==']
Decrypted file identifiers:
[b'Nepal,0', b'Nepal,986', b'Nepal,1044']
Time required to search 0.00608372688293457

In the above segment the search query "Nepal" was mistyped as "Nepam". The algorithm first computes edit distance with the closest keyword in the list (Nepal) and returns the file identifiers associated with the keyword. Encryption and Decryption is done with Fernat's theorem.

4. Assessing the information loss

Most anonymization and pseudonymization techniques involve suppressing or reducing the level of details provided by the input information. A significant challenge for the security team and statistician is striking a suitable trade-off between the levels of details lost and secured data. The final aim is to reduce the disclosure risk and minimize the loss at the same time.

To calculate and assess the information loss for textual data anonymization, the proposed work uses mean square error, mean absolute and cosine

Table 3 Information loss values for textual data masked with fake texts

MSE	MAE	CS
22.359	4.232	0.714

Table 4 Information Loss Values for Textual Data masked with hashed data

MSE	MAE	CS
9.466	2.6271	0.772

similarity. The metrics mentioned (Tables 3 and 4) are preferred for calculating the overall information loss for continuous data.

Mean Squared Error (MSE)

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2 \tag{2}$$

Mean Absolute Error (MAE)

$$\frac{1}{n}\sum_{i=1}^{n}|Y_i - X_i| \tag{3}$$

Cosine Similarity (CS)

$$\frac{\sum_{i=1}^{n} A_i x\, B_i}{(\sum_{i=1}^{n} A_i^2) x (\sum_{i=1}^{n} B_i^2)} \tag{4}$$

Where,

- n = number of data points
- Y = Vector representation of Masked/Hashed Text
- X = Vector representation of original text
- A/B are documents which are compared

The proposed model achieves low MSE and MAE scores for the anonymization task using Fake data. The scores indicate similarity with the actual textual data and a low loss of details during the anonymization task. A high cosine similarity score is achieved, indicating a high similarity between the actual and anonymized texts.

The proposed model achieves lower MSE and MAE scores for the pseudonymization task using the hash of the sensitive data. The scores indicate similarity with the actual textual data and a low loss of details during the anonymization task. An even higher cosine similarity score is achieved, indicating a high similarity between the actual and anonymized texts.

4.2 Discussion

The results obtained indicate the efficiency of the proposed methodology for textual classification tasks on private and sensitive data.

The degree to which change in the datasets offered by differential privacy is controlled by the ϵ parameter, which establishes a limit on the change in the likelihood of any given outcome. It measures the privacy loss owing to differential changes in data. Accuracy is the measure of closeness of substituted output with the actual output. The lower epsilon value guarantees complete privacy (more noise) but fails to give the correspondingly high model performance scores. The elevated noise introduced to protect and hide the private information fails to ensure any good semantic coherence between individual components of the data in the textual corpus, reducing data utility in the process. The algorithm allows us to set a privacy budget and spend it as required and necessary through various stages of text processing.

The anonymization techniques adapted perform much better on the data. These privacy-preserving data mining methods aim to modify the original data so that the private contents of a user or a group remains screened post-mining. The performance scores obtained using the classification algorithms indicated a comparable performance with the models trained without any masking or interference. The technique used in the proposed work ensures that the data's private information is masked with the appropriate security token to ensure that the logical coherence and overall sentiment of the textual sentence are retained. The retained logical coherence in the sentence is why the model performs just as well as the model trained on an unprotected database. The algorithm strategically only hides the specific components of the textual data, which, if left exposed, could lead to unavoidable data breaches.

The proposed work preferred to employ replacement and pseudo-anonymization techniques over aggregated anonymization methods like K-Anonymity and L-Diversity due to their relative simplicity and ease of implementation in a textual context. The existing numerical generalization and categorical ID-based approach are very effective to standard databases. These approaches prove ineffective in a distributed architecture like OSN, wherein concrete values are often not present straightforwardly. Most of the work in the proposed domain involves the protection of user friends and their location. Textual information via tweets and profile attributes has an abundance of sensitive information which can even reveal identities. Our work deals with anonymizing the private content before data publishing.

Cryptographic mechanisms, although most effective, increase the complexity due to overhead in key generation and management. Existing randomization and perturbation methods involve the addition, deletion or swapping of nodes and edges, which compromises the data utility and its integrity if applied to sensitive attribute protection. A severe drawback observed in these earlier methods is the loss in data quality post anonymization. The proposed study, unlike its earlier counterparts, focuses on multiple categories of sensitive attributes. Traditional methods have usually applied differential privacy for either node or edge protection only. Our approach, however, uses the same to conceal sensitive information. The proposed work achieves a high logical similarity for the anonymized data compared to the actual text. The high cosine similarity and lower values of mean squared error and mean absolute error indicate that work successfully captures the details of the text post masking. The work maintains high classification model scores while parallelly reducing the risk while communicating and data and retaining the overall detail of the data.

At the server end, the proposed modified searchable encryption has proven merit in scalability and compatibility with our model while providing suitable utility and integrity. It shows a clear improvement over traditional keyword-based searchable encryption methods by being more tolerant of minor typos without compromising privacy.

5 Conclusion and Future Works

The proposed approach facilitates a more controlled means for data dissemination in a public OSN. Predicting trust in OSN is a daunting task due to the lack of physical connectivity and complete factual information. A selective inconspicuousness method is proposed. It first identifies the personally identifiable information from tweets and user bio, anonymizes it and transfers the crucial data via the cloud, from where authorized users can retrieve them via proposed searchable encryption.

Some of the open research problems which can be a possible scope for future work research are listed in the subsequent paragraph. Steps to reduce algorithmic complexity, computation overhead and enforce seamless key exchange in selective encryption can be an appropriate undertaking for the future. The current study is only limited to sensitive data protection in social media. Anonymizing sensitive links and group memberships was beyond the scope of our work. Devising a way to cater to the capriciousness of these media giants effectively can be a good scope for future work in this

domain. Noise addition in differential privacy can be made more efficient by considering parameters like the level of privacy needed.

Further investigations are suggested in anonymizing interrelated heterogeneous data like age and occupation. Research can be done to incorporate computationally efficient access based deanonymization to promote studies in pseudonymization. Search results of the proposed searchable encryption technique can be made more robust by incorporating semantic analysis and natural language processing, and suitable ranking mechanisms.

References

[1] E. Raad and R. Chbeir, "Privacy in Online Social Networks," in Security and Privacy Preserving in Social Networks. Springer-Verlag Wien, 2013, pp. 3–45. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00975998

[2] J. Gehrke, E. Lui, and R. Pass, "Towards privacy for social networks: A zero-knowledge based definition of privacy," in Theory of Cryptography, Y. Ishai, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 432–449.

[3] Senthil Kumar N, Saravanakumar K, and Deepa K, "On privacy and security in social media – a comprehensive study," Procedia Computer Science, vol. 78, pp. 114–119, 2016, 1st International Conference on Information Security Privacy 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050916000211.

[4] H. Krasnova, O. Günther, S. Spiekermann, and K. Koroleva, "Privacy concerns and identity in online social networks," Identity in the Information Society, vol. 2, no. 1, pp. 39–63, Dec 2009. [Online]. Available: https://doi.org/10.1007/s12394-009-0019-1

[5] A. Srivastava, "Enhancing Privacy in Online Social Networks using Data Analysis," Birla Institute of Technology and Science, Pilani, Tech. Rep., 2015.

[6] S. Ali, N. Islam, A. Rauf, I. U. Din, M. Guizani, and J. J. P. C. Rodrigues, "Privacy and security issues in online social networks," Future Internet, vol. 10, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1999-5903/10/12/114.

[7] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," IEEE Communications Surveys Tutorials, vol. 16, no. 4, pp. 2019–2036, 2014.

[8] H. AbdulKader, E. ElAbd, and W. Ead, "Protecting online social networks profiles by hiding sensitive data attributes," Procedia Computer Science, vol. 82, pp. 20–27, 2016, 4th Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916300187.

[9] A. Pawar, S. Ahirrao and P. P. Churi, "Anonymization Techniques for Protecting Privacy: A Survey",2018 IEEE Punecon, 2018, pp. 1–6, doi: 10.1109/PUNECON.2018.8745425.

[10] K. Hasanzadeh, A. Kajosaari, D. Häggman, and M. Kyttä, "A context sensitive approach to anonymizing public participation GIS data: From development to the assessment of anonymization effects on data quality," Computers, Environment and Urban Systems, vol. 83, p. 101513, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0198971520302465.

[11] N. Lisin and S. Zapechnikov, "Methods and approaches for privacy-preserving machine learning," Advanced Technologies in Robotics and Intelligent Systems Mechanisms and Machine Science, pp. 141–148, 2020.

[12] M. Gaur, "Privacy preserving machine learning challenges and solution approach for training data in erp systems," 2020.

[13] M. Siddula, Y. Li, X. Cheng, Z. Tian, and Z. Cai, "Anonymization in online social networks based on enhanced equi-cardinal clustering," IEEE Transactions on Computational Social Systems, vol. 6, no. 4, pp. 809–820, 2019.

[14] A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data," Journal of King Saud University – Computer and Information Sciences, vol. 31, no. 4, pp. 426–435, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157817304093.

[15] W. Ouyang and Q. Huang, "A privacy preserving algorithm for mining rare association rules by homomorphic encryption," in 2019 6th International Conference on Systems and Informatics (ICSAI), 2019, pp. 1403–1407.

[16] M. Dias, A. Abad, and I. Trancoso, "Exploring hashing and cryptonet based approaches for privacy-preserving speech emotion recognition," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 2057–2061.

[17] W. Wei, S. Liu, W. Li, and D. Du, "Fractal intelligent privacy protection in online social network using attribute-based encryption schemes," IEEE Transactions on Computational Social Systems, vol. 5, no. 3, pp. 736–747, 2018.

[18] K. R. Macwan and S. J. Patel, "k-NMF anonymization in social network data publishing," The Computer Journal, vol. 61, no. 4, pp. 601–613, 2018.

[19] M. Yuan, L. Chen, P. S. Yu, and T. Yu, "Protecting sensitive labels in social network data anonymization," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 3, pp. 633–647, 2013.

[20] H. Huang, D. Zhang, F. Xiao, K. Wang, J. Gu, and R. Wang, "Privacy-preserving approach PBCN in social network with differential privacy," IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 931–945, 2020.

[21] N. Wu, F. Farokhi, D. Smith, and M. A. Kaafar, "The value of collaboration in convex machine learning with differential privacy," 2019.

[22] N. Holohan, S. Braghin, P. M. Aonghusa, and K. Levacher, "Diffprivlib: The ibm differential privacy library," 2019.

[23] A. Triastcyn and B. Faltings, "Bayesian differential privacy for machine learning," 2020.

[24] D. Xu, S. Yuan, and X. Wu, "Achieving differential privacy and fairness in logistic regression," in Companion Proceedings of the 2019 World Wide Web Conference, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 594–599. [Online]. Available: https://doi.org/10.1145/3308560.3317584.

[25] L. Chen, K. Huang, M. Manulis, and V. Sekar, "Password-authenticated searchable encryption," International Journal of Information Security, 2020.

[26] M. A. M. Ahsan, M. Y. Idna Bin Idris, A. W. Bin Abdul Wahab, I. Ali, N. Khan, M. A. Al-Garwi, and A. U. Rahman, "Searching on encrypted e-data using random searchable encryption (ranscrypt) scheme," Symmetry, vol. 10, no. 5, 2018. [Online]. Available: https://www.mdpi.com/2073-8994/10/5/161.

[27] Roesslein, J. (2020). Tweepy: Twitter for Python! URL: Https://Github.Com/Tweepy/Tweepy.

[28] "Faker Is a Python Package That Generates Fake Data for You." PythonRepo, https://pythonrepo.com/repo/joke2k-faker-python-testing-code bases-and-generating-test-data.

[29] Yuanxin Li, Darina Saxunová, "A perspective on categorizing Personal and Sensitive Data and the analysis of practical protection regulations", Procedia Computer Science, vol. 170, 2020, pp. 1110–1115, ISSN 1877-0509, Available: https://doi.org/10.1016/j.procs.2020.03.060.
[30] Hutto, C.J. and Gilbert, E.E. (2014). "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.", Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
[31] A. Srivastava and G. Geethakumari, "Measuring privacy leaks in online social networks," in 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013, pp. 2095–2100.
[32] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in 2010 Proceedings IEEE INFOCOM, 2010, pp. 1–5.

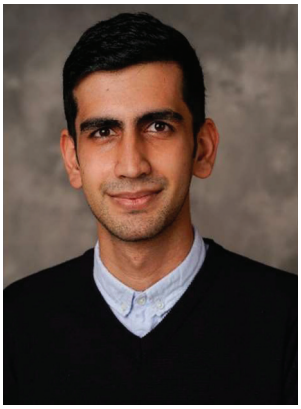
Biographies

Nisha P. Shetty has acquired her bachelor and master's degree from Visvesvaraya Technological University. She is currently pursuing her doctorate at Manipal Institute of Technology, Manipal. She is working in the area of social network security.

Balachandra Muniyal's research area includes Network Security, Algorithms, and Operating systems. He has more than 30 publications in national and international conferences/journals. Currently he is working as the Professor in the Dept. of Information & Communication Technology, Manipal Institute of Technology, Manipal. He has around 25 years of teaching experience in various Institutes.

Niraj Yagnik pursued his bachelor's degree at Manipal Institute of Technology, Manipal – India. His areas of interest include Data Science and Natural Language Processing. He is currently working as a Senior Developer at ICICI Lombard.

Tulika Banerjee pursued her bachelor's degree at Manipal Institute of Technology, Manipal – India. Her areas of interest include Data Science and Natural Language Processing. She is currently working as a Software Engineer at Amadeus Labs.



Angad Singh pursued his bachelor's degree at Manipal Institute of Technology, Manipal – India. His areas of interest include product management, market & growth strategy, customer analytics, and community building.