Revitalizing Network Privacy Preservation with Genetic Algorithms

[1]M.Nagaraju Naik. [2]Prof.M. Padmavathamma
[1]Research Schlaor, [2]Research Supervisor
[1,2]Department of Computer Science
S V University, Tirupati

Abstract : An increased use of data driven applications and integrated systems have caused an rushing expansion in data volumes and increase in the number of digital records, over the past few decades. Exponentially growing data volumes being processed by large-scale distributed data-intensive applications have placed an increasing pressure on the underlying storage services for timely and efficient storage and retrieval of the data. The use of cloud storage is among the best strategies to efficiently store growing volumes of data. However, outsourcing data to public cloud storage leads to the challenge of data privacy preservation. The paper talks about using Genetic Algorithm (GA) in Network Security. GA is one of the commonly used approaches on data mining. The paper puts forward a GA approach for classification problems. The proposed method uses a merger of privacy-aware selective cryptographic techniques, Genetic Algorithm (GA) and deep CNN. The experiments and comparative analysis depict that our proposed method outperforms others under consideration, in terms of accuracy, precision, recall and F1- score respectively.

I. INTRODUCTION

The genetic algorithm (GA) is the best-known optimization method that is centred on the environment. In GA, the search for the solution space mimics the natural procedure that happens in nature is considered. GA considers, population to be individuals; all, termed a chromosome, signifies the ideal solution to the issue. The issue being resolved is distinct using the objective function. Based on how "good" the provided individual is fixed to the objective function. The value that indicates an individual's quality is assigned to it depending on how well it fits the objective function. This value, which is also a key evaluation aspect, is known as individual fitness. Highly regarded people are more likely to be chosen for the next generation of the population. Three operators are used in GAs: selection (which creates a new population of individuals centred on the fitness values of individuals from the preceding generation), crossover (which typically involves the exchange of individual parts between two individuals selected for the crossover), and mutation (which involves the random alteration of the values of specific genes). GA are widely used to create superior solutions for the optimization and search problems. Genetic algorithms run in parallel on nodes or multiple processors, letting them to process huge sums of data rapidly and effectively. In this study, *GA* is applied to fix privacy concerns and *data security*. The algorithm of GA is shown in Algorithm 1.

Algorithm 1 - Genetic Algorithm

Input: Input data
Output: Best Feature Extracted from Data

1. $Determine\ Objective\ Function\ (OF_i)$
2. $Assign\ Number\ Of\ Generation\ to\ 0(t_i = 0)$
3. $Randomly\ Create\ Individuals\ in\ initial\ \ popoulation\ P_i(t_i)$
4. $Evalute\ Individuals\ in\ PopulationP_i(t_i)\ using\ (OF_i)$
5. $While\ Termination\ Criterion\ is\ not\ satisfied\ do$
6. $t_i = t_i + 1$
7. $select\ the\ individuals\ to\ population\ P_i(t_i)\ From\ P_i(t_i - 1)$
8. $Change\ \ individuals\ of\ P_i(t_i)\ \ Using\ Crossover\ and\ Mutation$
9. $Evalute\ \ individuals\ in\ population\ \ of\ P_i(t_i)\ \ Using\ OF$
10. $End\ While$
11. *Return Signal*

A greedy algorithm is a method for resolving a problem by choosing the best choice available at the instant. This technique is employed to solve optimisation issues. An optimization issue is an issue that claims either maximum or minimum results. The greedy method is a straightforward and modest approach. Based on the information present currently, a decision is taken and this is the chief function of its technique. Regardless of any current data present, the decision is taken without perturbing about the impact of the current decision in the near future. The key advantage of this algorithm is that it is simple to understand and implement. Usually, this technique is employed to decide the possible solution that may not or may be ideal. The possible solution is a subgroup that fulfils the provided standards. The possible solution is the solution that is superior and most preferable solution in the subgroup. In the event of possible, if few solutions fulfil the given standards, then such solution are considered possible, while the ideal solution is the superior solution among all solutions. The greedy algorithm is shown in Algorithm 2.

Algorithm 2 - Greedy Algorithm

$$Parameters(S_{ii}) - Set\ of\ blocks$$
$$Output\ - Superdata\ of\ set\ S_{ii}$$
$$While\ \left||S_{ii}|\right| > 1$$
$$do \begin{cases} choose\ s_{ii_1}, s_{ii_2} \in S_{ii}\ Such\ that\ overlap\ (s_{ii_1}, S_{ii_2}) is\ maximal \\ S_{ii} \leftarrow (S_{ii} \setminus \{s_{ii_1}, S_{ii_2}\}) \cup \{merge(s_{ii_1}, S_{ii_2})\} \end{cases}$$
$$return\ (remaining\ data\ in\ S_{ii})$$

Greedy algorithms are employed for crossovers, although these must be randomized to offer better outcomes. Special purpose enhancing mutations improve domain search too large for conventional mutation to be valuable. The greedy genetic algorithm yields schedule typically within a limited duty of the ideal solution. The greedy based algorithm is shown in Algorithm 3.

Algorithm 3 - Greedy based Genetic Algorithm*:*

$$Parameters(S_{ii}) - Set\ of\ blocks$$
$$Output\ - Superdata\ of\ set\ S_{ii}$$
$$While\ \left||S_{ii}|\right| > 1$$
$$do \begin{cases} choose\ s_{ii_1}, s_{ii_2} \in S_{ii}\ Such\ that\ overlap\ (s_{ii_1}, S_{ii_2}) is\ maximal \\ S_{ii} \leftarrow (S_{ii} \setminus \{s_{ii_1}, S_{ii_2}\}) \cup \{merge(s_{ii_1}, S_{ii_2})\} \end{cases}$$
$$return\ (remaining\ data\ in\ S_{ii})$$
$$Begin$$
$$Generate\ the\ initial\ population\ POP_{ii}(0);$$
$$Estimate\ POP_{ii}(0);$$
$$Repeat$$
$$Select\ parents;$$
$$Generate\ new\ chromosomes\ using\ crossover;$$
$$Applied\ mutation\ on\ the\ new\ chromosomes;$$
$$Applied\ greedy\ sequential\ function;$$
$$Estimate\ POP_{ii}(t_i);$$
$$Until\ (Terminating\ condition\ is\ reached);$$
$$End;$$

As shown in Figure 1. GA initiates with an unsystematically created initial chromosome population, in that every chromosome signifies a latent solution to the provided problem. Every chromosome has a fitness value assigned to it. And that quantifies how well a certain problem has been solved. Using evolutionary operators like selection, mutation, and crossover, the population evolves each generation in the direction of greater fitness. Until a solution is discovered or the maximum number of iterations has been reached, this process continues. The conventional genetic algorithm's initial population was chosen using a time-consuming and unstable random method. The enhanced genetic algorithm uses a greedy strategy instead, reducing these drawbacks and accelerating the rate at which the initial population meets the ideal.

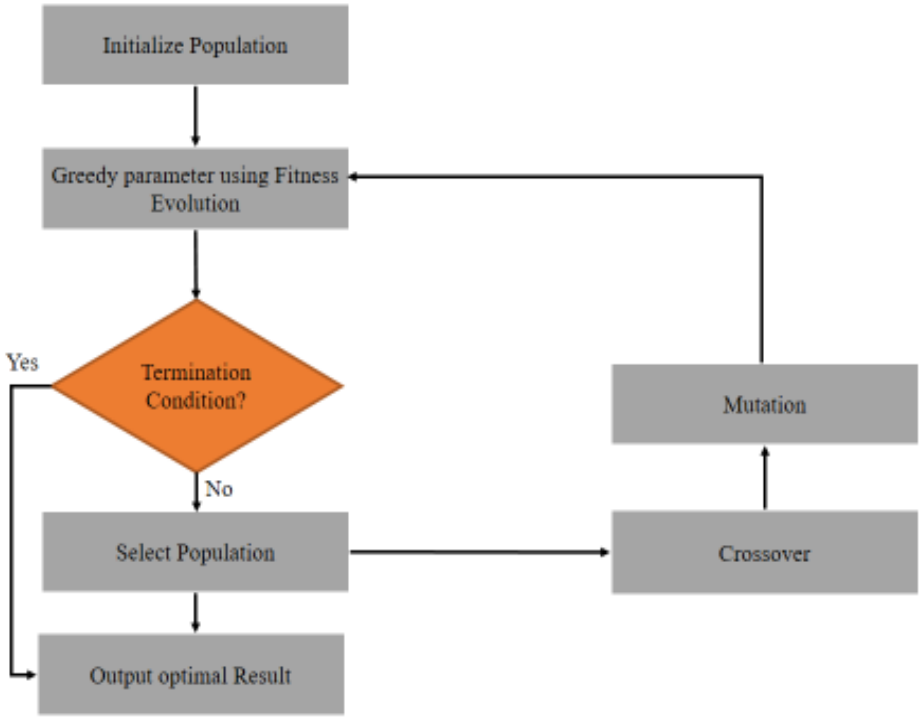

Figure 1. Greedy based genetic algorithm architecture

III. METHODOLOGY

Initially, the Cybersecurity KddCup99 dataset is loaded for the research purpose. Once, the dataset is loaded it is then pre-processed. The pre-processing of the data involves the various features such as filling the missing values and column reduction and removing the unnecessary noise. During the pre-processing stage, the feature scaling takes place which standardizes the individual features of the data. Once the data is pre-processed then it performs the encryption process for the multi keyword ranked search using the hybrid Elliptical Curve Cryptography (ECC) and Advanced Encryption Standard (AES) algorithm. The keywords with the highest significance will be stored as the partition. The documents corresponding to keywords are stored in the same partition. In this way, various partitions will be found along with the keyword and its corresponding documents. Finally, the documents will be displayed while typing the respective keywords. The documents are encrypted using the encryption algorithm to safeguard the confidential data. The admin can directly read the decrypted data meanwhile, the other users need to register and then can access the decrypted data.

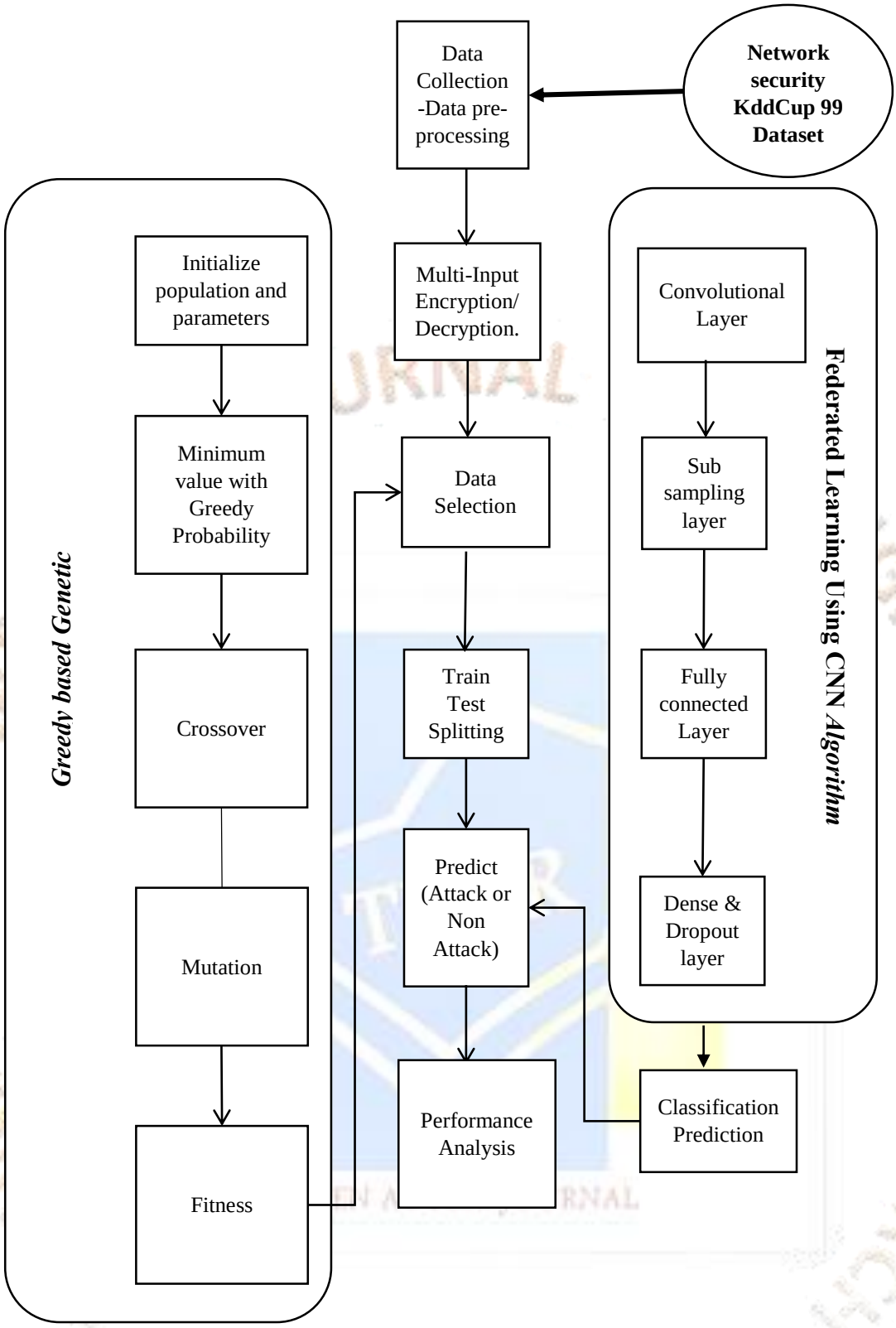

Figure 2. The overall flow for the study

In order to overcome the overfitting problem, the feature extraction process is performed by the greedy based genetic algorithm. The genetic algorithm performs the inheritance, mutation, and crossover and selection operation and also incorporates several greedy principles; hence it is termed as the greedy based genetic algorithm. The time complexity is reduced by using this greedy based genetic algorithm. The basic steps involved inside the greedy based genetic algorithm are,

- At the first stage, the generation of the initial population and parameters takes place. The initial population is produced in a random manner.
- The minimum value will be selected on the basis of locally optimum solution by means of the greedy approach.
- Then the mutation and crossover operations are performed correspondingly.

- The higher the fitness value, the higher the chances for selection of the data.

III. EXPERIMENTAL RESULTS

The experimental outcomes of the proposed method presented in figure 2 and in Table 1 confess that the proposed method is effective in achieving the accuracy, precision, recall rates and F1-score.

Table 1: Performance metrics and values of proposed model

Metrics	Accuracy	Precision	recall	F1-score
Values	99.3	96	96	96

Figure 2. Performance metrics of the proposed model

IV. Conclusion

Feature extraction process using the greedy based genetic algorithm improves the prediction accuracy of the proposed method.

References

[1] Mishra, T. S. Jabar, Y. I. Alzoubi, K. N. J. C. Mishra, C. Practice, and Experience, "Enhancing privacy-preserving mechanisms in Cloud storage: A novel conceptual framework," p. e7831, 2023.

[2] varshachoudhary, "Cyber Security Policy," 2022.

[3] K. Kumar, V. Kumar, and Seema, "Security and Privacy Preservation for Data Communication Network," 2022.

[4] T. Øren and S. P. Fosser, "Multi-Cloud Information Security Policy Development," University of Agder, 2023.

[5] Osano. (2023). *5 Emerging Data Privacy Trends in 2023*. Available: https://www.osano.com/articles/data-privacy-trends

[6] D. p. manager. (2023). *5 things you need to know about Data Privacy [Definition & Comparison]*. Available: https://dataprivacymanager.net/5-things-you-need-to-know-about-data-privacy/

[7] U. A. Shah, M. Hussain, M. Saddiqa, and M. S. J. L. P. P. Yar, "Problems and challenges in the preservation of digital contents: an analytical study," pp. 1-12, 2021.

[8] E. Bertino, I. N. Fovino, L. P. J. D. M. Provenza, and K. Discovery, "A framework for evaluating privacy preserving data mining algorithms," vol. 11, pp. 121-154, 2005.

[9] P. Yang, N. Xiong, and J. J. I. A. Ren, "Data security and privacy protection for cloud storage: A survey," vol. 8, pp. 131723-131740, 2020.

[10] I. Sudha, R. J. I. J. o. M. Nedunchelian, Simulation,, and S. Computing, "A secure data protection technique for healthcare data in the cloud using homomorphic encryption and Jaya–Whale optimization algorithm," vol. 10, no. 06, p. 1950040, 2019.

[11] S. M. J. I. A. Kasongo, "An advanced intrusion detection system for IIoT based on GA and tree based algorithms," vol. 9, pp. 113199-113212, 2021.

[12] C. Dhasarathan *et al.*, "COVID-19 health data analysis and personal data preserving: A homomorphic privacy enforcement approach," vol. 199, pp. 87-97, 2023.

[13] O.-A. Kwabena, Z. Qin, T. Zhuang, and Z. J. I. A. Qin, "Mscryptonet: Multi-scheme privacy-preserving deep learning in cloud computing," vol. 7, pp. 29344-29354, 2019.