Cluster Computing
https://doi.org/10.1007/s10586-022-03667-3

Multi-objective optimization-based privacy in data mining

Hemanta Kumar Bhuyan[1] · Vinayakumar Ravi[2] · M. Srikanth Yadav[1]

Received: 5 December 2021 / Revised: 17 April 2022 / Accepted: 21 June 2022

Abstract
This paper addresses the data privacy based on interactive computation using an optimization model in data mining. When data are computed or sharing among users in online, it needs to maintain privacy for all computation during sharing of data. But user choice-based privacy is not available when sharing of data is required for data mining computation which is a big challenge for data privacy. Thus, we proposed the framework for anonymity of data privacy using various methods of multi-objective models as per the requirement of privacy. The proposed framework is designed with the help of two objects such as computational cost and privacy based on optimization model. Our framework maintains the balance between above objects as per user demands, i.e., increasing the privacy with decreasing the computational cost. In this model, the domain of privacy and computational cost for optimization problem solves the entity privacy requirements in a computing environment. We have used various methods such as Gaussian and uniform distribution, confidence interval, activation function, linear membership function with distinguish manner for maintaining of privacy and cost. As per the uniform distribution and parameter α-cut value for noise data, the optimal value is made accordingly. Example: for $\alpha = 0.2$, and uniform distribution $(-1, 1)$, the optimal value is 0.0058. Similarly, as per different α values, classifiers result is different like $\alpha = 0.2$ and 0.4, Multilayer perceptron values are 4.01 and 1.61 respectively. The solution of the proposed model controls the amount of privacy with complete freedom of choice of users with utmost flexibility.

Keywords Multi-objective optimization · Privacy · Uniform distribution · Active function · Data mining

1 Introduction

Securing computation in privacy model has been issued in privacy preserving data mining (PPDM). In PPDM, data owners and data miners are obliged to cooperate in good faith while releasing private data. Different authors have developed their own privacy model using various approaches. For different data transmission, various network models have been explained as per their own methodology such as Harris Hawks optimization [1], optimization with differential evolutionary model [2], peer to peer wireless mesh networks [3] etc. The limitation of existing work as: the monolithic owner does not make privacy decisions at the level of data processing. Both data owner and data miner are under cooperative in nature when the data are released by data owner for data mining task. But there is a chance for leakage of data during data sharing for data mining task. To avoid leakage of data during data sharing for data mining task, data owner wants to maintain privacy when it needs to share its data. Under these circumstances, the data owner does not want to release his data without maintaining own privacy Thus, it is a challenging task in data mining during sharing of data-by-data owner.

The data protection differs in computing environments, depending on personal details in the mining process. During mining process of data or sharing of data, there may be issues of individual privacy for data owners when working with several resource of data. Thus, data mining task need to take into account of individual privacy to be effective.

✉ Vinayakumar Ravi
vravi@pmu.edu.sa

Hemanta Kumar Bhuyan
hmb.bhuyan@gmail.com

M. Srikanth Yadav
srikanthyadav.m@gmail.com

[1] Department of Information Technology, Vignan's Foundation for Science, Technology and Research (Deemed to Be University), Guntur, Andhra Pradesh, India

[2] Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia

Published online: 12 July 2022

Thus, we proposed multi-objective approach for solving personal data mining problems when data owners want to release their data for mining task. We have considered two objectives such as privacy and computational cost to design multi-objective optimization model, where data owner can release their data as their own choice of privacy. Above two objects are maintained with certain balancing of evaluation where data owner gets benefits as his own choice of level of objects. Our proposed model tries to maintain the level of both objectives like maximize the privacy with minimize the computational cost.

Our proposed multi-objective optimization model finds out the solution using constraints which are expressed through privacy and computational cost. Since maximize the privacy and minimize the cost is another challenging task through optimization model, so we have considered different statistical approaches to solve this problem. We use linear combination of utility factors of privacy and computational cost and its corresponding measurements to develop the optimization model. Utility components also added to balance the optimization model. Constraints are designed through certain interval with the help of statistical approach i.e., uniform distribution. Quantification privacy is considered to define various level privacy for maintaining strong privacy during processing of data with the help of Gaussian uniform approach. When different constraints are designed, we considered α-cut constraints to find out certain level measurements for both privacy and computational cost. Thus, optimized performance is created accordingly with maintaining maximize privacy and minimize computational cost. As per the proposed model, the experimental performance is considered with various level of objects. Apart from above model, we have used various methods such as Gaussian and uniform distribution, confidence interval, activation function, linear membership function with distinguished manner for maintaining of privacy and cost. As per the uniform distribution and parameter α-cut value for noise data, the optimal value is made accordingly. Although we use various methods and evaluate its corresponding methods, still, here we explained the major contribution of this paper as follows.

(a) The multi-objective optimization model has developed based on two objects such as privacy and computational cost.
(b) The decision variables are designed through utility factors with its corresponding measurements.
(c) Individual utility components are balanced and tried to maintain maximum privacy and minimum computational cost.
(d) Measurement of privacy is developed through confidence interval which help to find out different level of privacy.
(e) We used α-cut constraints to find out various optimal test where data owner can choose his privacy as per best optimal value.

The remainder of the paper has been assembled in the order described below. A literature survey has explained to the development of the proposed model in Sect. 2 as related work. The problem statements are explained in Sect. 3. We develop the optimization model using privacy and computational framework using various constraints issues in Sect. 4. In Sect. 5, we used optimization constraints to define our best possible model. Section 6 describes the experiment evaluation as per proposed model and analyzed on experimental results. The brief description is concluded before ending the paper in Sect. 7.

2 Related work

The privacy requirement is a major part in data mining, in which the data owner and the data miner used the data for the sake of sharing. Several privacy-preserving approaches are enabled for data mining analysis. As per PPDM utility-based study, the various factors have been investigated by different researchers. The optimization models through fuzzy have designed by [4, 5] and the privacy measurement is optimized to preserve the confidentiality of published results. Many experiments have been done with one of these ideas, as per Bayardo et al. [6]. The search algorithm has been bound to the yearning consistency of the result without any regard to efficiency like for each data record collection, the algorithm has to be reapplied, for the best possible anonymity, the model incorporates a ℓ-diversity strategy and lack of privacy details. Different privacy model are developed such as Gheisari et al. [7] developed Ontology-Based Privacy-Preserving in IoT-based smart city, where as Omar et al. [8] have used Blockchain and artificial intelligence-based privacy-preserving in healthcare system. Using privacy enhancement strategies for the most significant degree of anonymity, security, privacy, and productivity has been a big challenge. Different privacy models have been established around the world, including the Bayesian approach, k-anonymity, and the probability-based model of privacy. There are distinct templates for risks that help categorize distinct usages of privacy.

To effectively and efficiently mine rules, disruption of contact and computing is minimized in the P2P context, but different issues for data are made to many stakeholders. All are more or less available in a distributed fashion using a decentralized algorithm. However, lower connectivity costs and limited computational resource availability make coordination more costly for the end consumer [9].

Observations on the uses of social networks have raised concerns about user privacy. Differential privacy is implemented in PPDM to guard against intrusion and is defined in workload partitioning [10]. Differential privacy has been used in different ways in studies for different reasons. However, having been incited to defend individual privacy, PPDM has been supported in maintaining their data [11] to preserve their privacy.

For a variety of functional issues, limits, some versatility has been essential. Those various parties are allowed different degrees of control in the problem-solving process to provide the problems differing requirements with suitable solutions. The federated learning scheme is used for privacy preserving with data aggregation in [12]. Due to a lack of verifiable evidence, multi-objective optimization is thought to be a fuzzy challenge. The multi-criteria decision dilemma originates under a cloud of uncertainty, with classical optimization problem. The secure summation for privacy preserving data has been developed [13]. Few authors also developed fuzzy model for features data [14] in which it needs to consider many objectives. It can employ multi-objective optimization and dynamic decision making known as fuzzy decision making that evaluates a variety of degrees of options and a weighting of parameters [15]. Vague binary decision making is done using multi-criteria search techniques and the resolution in a linear programming problem [16]. Multi-criteria decision-making and intuitionistic methods are elaborated in linear programming. Some have equated the decision-making in supply chain management with the Pareto solution of the fuzzy optimization problems.

Boyd [18] and Deb [19] provide distinct dilemma definitions. A fuzzy optimization is a pragmatic approach to addressing the optimality problem; fuzziness will still be important when solving variable costs. Two types of constraint problems may be of fuzzy optimization, namely (a) ones with constraints such as linear and (b) unconstrained. The linear fuzzy solution has been addressed in [20]. In the non-linear, the questions can be approached in several ways. For further information on non-linear and linear programming issues, it can study [17] and [21], respectively. Very often, it reaches away until it can meet the constraints, and often, some problems arise as it becomes impossible to find a solution. Many times, non-dominated sets are thought to be ideal solutions to the dilemma. Non-dominated solution and non-dominated data sets are presented in [18, 19], where it is found that the detailed information on Pareto optimization is provided; therefore, the appropriate and more privacy-friendly data mining methods have been applied to several data mining issues. Privacy preserving model has been developed by Bhuyan et al. [23, 24, 28, 29] individually, whereas differential privacy, general privacy, crypto analysis-based privacy model has developed in [25–27]. Using the concepts of the Pareto principle in data mining, Kamila et al. [30] utilize the following strategies: To perform class-differentiation in data mining, the certainty of feature data must be controlled [31–33], But the privacy of the data is protected by a different approach, known as Quantifying Differential Privacy, and a related methodology known as Privacy-Preserving Graphs maintained accordingly. Further, different authors have also used privacy preserving model to analyze various public data for IoT based smart cities and agricultures as per requirements [34–36].

3 Problem statements

Data sharing by all parties is important in data mining task, where data owner want to preserve the privacy during sharing or computation his data. Based on privacy preservation of data, the computing cost will differ when data computed by data owners. According to the owner's data protection requirements, various approaches can maintain privacy without violating confidentiality. But the data miner is still searching the methods to maintain privacy when data is being published. For secrecy of data, privacy calculation is critical when making decisions on the acceptable level of privacy for each owner of each record. Although the privacy expense cannot be ignored entirely, still the privacy with computational cost can be considered for data computation. In this instance, data privacy may be optimized by paying the least amount of computing costs to the data owner. Since the data owner has a broader goal, our model uses a multi-objective approach to construct a conceptual model which satisfies the minimal data owner confidentiality.

Generally, data owners want different levels of privacy concerns, and the above descriptions are not helpful for any of these types of data. Uncertainty over the cost of optimal privacy and maximum usefulness are often poorly described. Thus, it needs that the data owner's privacy and costs are equally essential for solving the optimization issue. The multi-objective optimization problem is considered to create the following mathematical objects (i.e., privacy and cost). The model is formulated as a multiparty distributed application because each group contributes their data to protect their privacy. Regarding all of the issues mentioned above, we ought to respect the privacy of each party as a data contributor.

A multi-object optimization model is now established with various solution of the optimization problem. An optimization technique called multi-objective is constructed from many diverging priorities. It can be represented mathematically as Eq. (1) with corresponding constraints:

$$\begin{aligned} &\text{Max } f(x) = [f_1(x), \ldots, f_M(x)]^T \\ &\quad \text{Subject to } P_j(x) \le g, && \forall j = 1 \ldots P \\ &\quad\quad Q_k(x) = h, && \forall k = 1 \ldots q \\ &\quad\quad x_i^l < x_i \le x_i^u && \forall i = 1 \ldots m \end{aligned} \tag{1}$$

where $f_1,\ldots, f_M$ are M scalar objectives with f_i: $R^m \rightarrow R$, P_j and Q_k are mapping $R^m \rightarrow R$ for both constraint functions, g, h ε R, and variables are precisely bounded between x_i^l and x_i^u. The solution is considered as a vector $x' = \{x'_1, x'_2,\ldots, x'_m\} \in R^m$ for above multi-objective optimization problem.

We treat the model as if it is the case that the data miner does optimization on the model. Thus, it is created to serve the purpose (a) privacy needs, (b) minimize computing cost. Functionality and complexity of data are used to calculate data miner-driven design optimization. For the model to yield the results, we need the following decision variable constraints to be developed.

3.1 Decision variable constraints

The following two decision variables are developed for proposed model. These decision variables paly the important role throughout the whole paper. Our multi-objective optimization model is designed using following two variables which is helped to data owner to decide how much privacy can maintain as per computational cost.

(a) I^{HP} = Length of interval for required privacy, i.e., high privacy (HP) or low privacy (LP).
(b) I^{CC} = Computational cost, i.e., high priority-based objective with low computational cost.

In this model, each data owner decides on their strategy and arbitrary behaviors to get the best scores. Both privacy and computational cost considered with dimensions of PPDM problem for each data holder. According to the dimensions of utility function $U_i(x)$, Eq. (1) may be recast as Eq. (2). The following vectors are considered for various dimensions of utility into multi-objective optimization as Eq. (2):

$$\text{Max } U_i(x) = \left[\sum C_{HP}U_{HP}, \sum C_{CC}U_{CC}\right] \tag{2}$$

with required constraints where usual notations have been defined successively. Again mathematically, we consider the weighted linear combination of the above dimensions of utility function for multi-objective optimization as Eq. (3),

$$\begin{aligned} &\text{Max } U_i(x) = a_i^{HP} \sum C_{HP}U_{HP} + \sum C_{CC}U_{CC} \\ &\text{Subject to} \quad \Sigma I_i^{HP}(x) \le g \\ &\quad\quad \Sigma I_i^{CC}(x) \le h \\ &\quad\quad \sum a_i^{HP} + \sum a_i^{CC} = 1 \\ &\quad\quad x_i^l \le x \le x_i^u \qquad \text{for all } i = 1, 2, \ldots n \end{aligned} \tag{3}$$

where (a) U_{HP}, U_{CC}, are utility factors of privacy and computational cost (b) C_{HP}, C_{CC} are the measurement computational cost (i.e., how much measured privacy and cost) of each utility factor, and also (c) a_i^{HP}, a_i^{CC} are weights of each utility component.

The privacy must be less than or equivalent to its predefined upper bound (in terms of time). There are two variables which connect to U(x) are associated with this space (a) and (b). The solution to this optimization issue is arrived by mutually binding restrictions on both the data owner and the data miner. The data owner determines the usefulness value by the data's privacy, and computing cost which are optimized separately by the data miner.

4 Privacy and computational cost framework

In this section, the privacy and computational measurements are considered to establish multi-objective optimization framework in different following subsections.

4.1 Quantification of privacy

Estimating each owner's privacy levels is accomplished by measuring how certain spans of privacy as their trust period. For deciding how much privacy fulfilling the degree of importance (α) where the trust interval is used. Recognize the initial data as private data is considered from each data owner. Thus, interval-based data privacy is developed by data miner where the owner believes in the accuracy of the results within certain interval. An interval of variation in trust is given below to explain the definition.

Definition 1 An interval is said to be a confidence interval if the estimation of this interval satisfies the level of significance (α).

For example, let the interval $[C_1, C_2]$ be generated by two constants C_1 and C_2 at the level of significance where the original data lie. We define the probability of confidence interval as Eq. (4).

$$P(C_1 < Z < C_2) = 1 - \alpha, \tag{4}$$

Here, Z is considered as a standard normal distribution which is determined by probability distribution with

confidence interval between C_1 and C_2. The C_1 and C_2 limits are referred to as certainty coefficients where the calculation needs the most confidence. For example, an alpha value of 0.5, 0.05, or 0.01 represents a 50%, 95% or a 99%confidence coefficient respectively in confidence level. This comes out to a proportional value of 0.674, or 1.96 or 2.58. Many of them use private data to talk about their confidence levels as follows.

Let Z = the value of the significant random variable, μ = the population mean, σ^2 = the square root of the sample size. The confidence interval is calculated using the Gaussian distribution as below.

Let a large random sample of size n is considered from large database with mean μ and variance σ^2, then the sample mean is $\bar{x} \sim N(\mu, \sigma^2/n)$, i.e., $Z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$. Using Gaussian distribution, the confidence interval is determined as Eq. (5).

$$P\left(\bar{x} - C\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + C\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \text{ for interval } [-C, C]. \tag{5}$$

Here the interval $\left[\bar{x} - C\frac{\sigma}{\sqrt{n}}, \bar{x} + C\frac{\sigma}{\sqrt{n}}\right]$ is called confidence interval using mean for individual data in the database. Individual data is considered as individual feature data in the database.

4.2 Privacy and cost optimization

Though computational cost is often derived from running the algorithms, it is always preferable to solicit optimal privacy. The aims for the design of the proposed model are discussed earlier. Using the multi-objective approach to handle privacy, we are able to derive a statistical model of privacy. Define with data privacy and cost as: (a) $f_{HP}(x)$: $R^m \rightarrow R$ and (b) $f_{CC}(x)$: $R^m \rightarrow R$ respectively. Here R is treated as a multi-dimensional input vector to calculate the most optimal value of privacy and cost as below. The optimization problem is further derived as Eq. (6):

$$\begin{aligned} &\text{Max } f(x) = [f_{HP}(x), f_{CC}(x)]^T \\ &\text{subject to} \quad x_i^l \le x_i \le x_i^u \quad \text{for all } i = 1 \ldots m \end{aligned} \tag{6}$$

where $x \in R^m$, and each x_i is bounded between x_i^l, and x_i^u for feasible solutions. Further, it can be reformulated the same optimization problem into a scalar optimal problem as Eq. (7),

$$\begin{aligned} &\text{Max } F = w^T f(x) = [w_1 f_{HP}(x) + w_2 f_{CC}(x)] \\ &\text{subject to} \quad x_i^l \le x_i \le x_i^u, \quad \forall i = 1 \ldots m \\ &\qquad w_1 + w_2 = 1 \\ &\qquad w_1, w_2 \ge 0 \end{aligned} \tag{7}$$

where w_1 and w_2 are the relative weights that a data miner uses for data privacy and cost, respectively. From the above methods, we derive the optimization model without constraints which is explained in Sect. 5.

5 Constraints for multi-objective optimization

When solving the solution of the optimization problem, random data is almost always involved. It is hard to provide accurate estimates of cost, but it is essential to provide options and evidence to decision-makers. In these instances, data sets are used to serve as a computational framework for multi-objective problems. The proposed model deals with multi-objective optimization that takes into account constraints data set. For specific multi-objective questions, the selection of non-dominated alternatives allows the right compromise on inter-dependent priorities. The model deals with multi-objective problems with no prior constraints. In addition, the various priorities inter-object dependencies lead to a new kind of multi-objective decision making [28, 32, 33]. Hence, the following incidents are to be relevant cases for discussion.

The data from different data owners have different characteristics, such as random, vague and fuzzy, which are used in the optimization problem. The demands of data owners are also fuzzy. In this paper, the proposed model deals with the fuzzy multi-objective optimization issue with fuzzy constraints. The complex fuzzy optimization problem is stated as Eq. (8).

$$\begin{aligned} &\text{Max } Y = cx \\ &\text{subject to } (Ax)_i \le b_i \text{ for all } i = 1, 2, \ldots m \\ &\qquad x_j \ge 0,\ x_j \in N,\ j = 1, 2, \ldots n \end{aligned} \tag{8}$$

where m, n represents a set of integer numbers, $c \in R^n$, $A = \sum_j a_{ij}$ and a_{ij}, $b_i \in R$.

Let the constraints defining the issue have a fuzzy nature where the decision maker/data miner is willing to permit any change ($\le^{DM}$) over restricted constraints as Eq. (9)

$$\begin{aligned} &\text{Max } Y = cx \\ &\text{subject to} \quad (Ax)_i \le^{DM} b_i, \quad i = 1, 2, \ldots, m \\ &\qquad x_j \ge 0,\ x_j \in N, \quad j = 1, 2, \ldots, n \end{aligned} \tag{9}$$

where $\le^{DM}$ refers to the data-mining conditional relationship.

However, some typical approaches are required to solve multi-objective decision-making (MODM) problems with supportive and conflicting objectives. It is very complicated to choose the optimal decision with an increasing number of objectives.

$$\text{Max}\{ < C_i, x \geq Y_i\} \tag{10}$$

under

$$x \in X = \{x \in R^n | Ax = b, x \geq 0, b \in R^m|\}$$

An innovative method is developed for solving MODM problems depending on the interdependences among the multiple objectives. Hence, the following cases are considered.

Let the optimization problem is defined on multiple objective function as Eq. (11).

$$\text{Max}\{f_1(x), f_2(x), \ldots, f_k(x)\}, x \in X \tag{11}$$

where f_i: $R^n \rightarrow R$, are objective functions, $x \varepsilon R^n$ is the variable and $X \subset R^n$.

Let a function

$$p_i(t) : R \rightarrow [0, 1]$$

where $p_i(t)$ determines the degree of decision maker's requirements on ith objective of value t. The degree of x in the data set with the help of p(x) as Eq. (12).

$$P_i(x) = p_i(Y(x)) \tag{12}$$

where $P_i(x)$ is considered as well compromise to solutions for ith objective. So, it is quite reasonable to search for a solution of the following auxiliary problem

$$\max_{x \in X}\{P_1(x), \ldots P_K(x)\} \tag{13}$$

where $P_i(x) \in [0,1]$. The Eq. (13) can be converted into a single objective problem

$$\max_{x \in X} T\{P_1(x), \ldots, P_K(x)\}$$

Now we consider active functions for proposed functions as Eq. (14)

$$p_i(t) = \begin{cases} 1 & if\ t \geq M_i \\ v_i(t) & if\ m_i \leq t \leq M_i \\ 0 & if\ t \leq m_i \end{cases} \tag{14}$$

where $m_i = \min\{Y_i(x) \mid x \varepsilon X\}$ and $M_i = \max\{Y_i(x) \mid x \varepsilon X\}$ with independent minimum and maximum of the ith objective and $v_i(t)$ is a function. For the linear membership functions, P_i is defined as Eq. (15)

$$P_i(x) = \begin{cases} 1 & if\ Y_i(x) \geq M_i \\ 1 - \dfrac{M_i - Y_i(x)}{M_i - m_i} & if\ m_i \leq Y_i(x) \leq M_i \\ 0 & if\ Y_i(x) \leq m_i \end{cases} \tag{15}$$

The objective functions are considered because it has many approaches to decide on alternatives that are closer to independent minima and maxima. As per our model requirements, we enable the values of Eq. (3) to adjust for the user's benefit (m_i is an ith goal). For purposes of linearity, the maximum and minimum deviations are defined,

$$m_i \leq \Sigma I_i^{HP}(x) \leq g_i + M_i \tag{16}$$

$$m_i \leq \Sigma I_i^{CC}(x) \leq h_i + M_i \tag{17}$$

The derivative of Eqs. (14) and (15) with expression (m_i and M_i) have been addressed as above Eqs. (16) and (17). Individual data owner can disclose personal data in return for ratification with undermining the desire of privacy. The above methodology can manage the linear constraints to overcome our problem for a better solution. Thus, the complete multi-objective optimization problem using α-cut constraints is as Eq. (18).

$$\begin{aligned} \text{Max } U_i(x) = a_i^{HP} \sum C_{HP}(U_{HP}) + a_i^{CC} \sum C_{CC}(U_{CC}) \\ \text{Subject to} \quad \Sigma I_i^{HP}(x) \leq g_i + (1-\alpha)d_i \\ \Sigma I_i^{CC}(x) \leq h_i + (1-\alpha)d_i \\ \sum a_i^{HP} + \sum a_i^{CC} = 1 \\ x_i^l \leq x_i \leq x_i^u \qquad \forall i = 1 \ldots n \end{aligned} \tag{18}$$

Here d_i and $(1 - \alpha)$ refers to users' needs vs. constraint requirements. Experiments may be conducted utilizing different conditions as per α values. Each data owner privacy is dependent on how often they deviate from it. The whole set of private data is accepted, and the interval tests it. For a high level of privacy it trusts, we have defined the utility value as a confidence interval. The utility factor is calculated by the execution of various data mining algorithms, which is simply a computational cost. The C_{HP} is the total cost and is calculated based on the confidence level of privacy that is supplied. A computing cost includes data processing and data cost; in other words, computation's total cost is influenced by both perturbed system execution. The two terms such as a_i^{HP} and a_i^{CC} are assumed as weight factors which satisfy $a_i^{HP} + a_i^{CC} = 1$. Thus $a_i^{HP} = 1 - a_i^{CC}$ or $a_i^{CC} = 1 - a_i^{HP}$ is the appropriate design of optimal test.

6 Experimental analysis

In an experiment, a multi-objective optimization problem is solved through concerning both privacy and computational cost when applied to experiments. We considered three classification approaches such as Naïve Bayes (NB), multi-layer perceptron (MLP), and classification and regression tree (CART) on real-world data for our proposed model.

6.1 Dataset

The adult data set is taken from UCI machine learning repository [22] and is considered for our experiments. This dataset contains 48,842 instances with 14 attributes (both categorical and integer). It is a census income dataset. The income source depends on its occupation and its education. When individual data disclose to public, he/she may be fallen in trouble by opponents.

6.2 Environments

Our proposed methods are experimented on a personal computer with an Intel core (2.92 GHz, 16.00 GB RAM, 64-bit OS, Windows 10 worked under Microsoft Office 2010), and also, we have used WEKA (Waikato Environment for knowledge analysis) data mining tools including Naïve Bayes classifier, Multilayer Perceptron, and CART on our dataset. It tries out three classifiers and find out the best classifier for most appropriate. The evaluation performs based on quantity of privacy measurements and computational cost.

6.3 Computational complexity analysis

We considered computational experiments using optimization model as per the data owner demand based on both privacy and computational cost. We used three individual privacy testing such as (a) individual privacy using optimization method (IPO) (proposed), (b) individual privacy using neural network (IPNN), (c) individual privacy using secure sum multiparty computation (IPSMC). When above approaches implemented as adult data set, the execution time is considered as shown in Table 1. The computational complexity and computational cost are analyzed as Tables 2 and 3 as above methods. Comparing other two methods, the computational performance of our proposed method is better than others. Thus, the individual privacy model using independent algorithms can be designed for data owner demands.

6.4 Experimental results

We used both real and randomized data to run the experiment for both privacy and computational costs. For perturbation, the random data is added in original data. Several collections of uniform data produce different types of perturbed data, according to the data owner demands. Consequently, the efficiency of a method increases from simple to complicated. With regards to privacy, it may also claim that computational cost is significant for high privacy. But an optimality test that has minimal computational cost is essential for model. For various perturbed data, the computational cost in Table 4 has changed.

When different noise data analyze through noise in format of {y, y/16, 16y, y/2, 2y}, then its corresponding computational cost values are generated as shown in Figs. 1, 2 and 3 as different classifiers. When compare to all above three figures, it is shown that MLP classifier get good results compare to other classifiers based on several perturbed data. In particularly, one point of noise 16y format is very high compare to other perturbed data through different noise using NB classifier performance.

We have generated several perturbed data based on {y/16, y/8, y/4, y/2, y, 2y, 4y, 8y and 16y} from Table 4 using two uniform noise distributions [0, 1] and [− 1, 1] and yields reliable results. Perturbed data provided detail, and in any case, put the original data x between [x − y, x + y], i.e., where y is uniform noise distribution and the length of privacy is 2y, i.e., the twice the noise distribution. If x lies between [x − y/2, x + y/2], then the length of privacy will be y. Again, if $x \in [x - y/n, x + y/n]$, then the length of privacy will be 2y/n. When $n \to \infty$ then $y/n \to 0$ and [x − y/n, x + y/n] is very close to x. In this case there is no requirement in adding or subtracting noise with original data. The data itself would not cause the noise to be added or subtracted. It is better to provide a cap on the amount of noise, and an individual exposed to an unlimited amount of valuable information distributed to them. With a reduction in noise value, the original data will become less privacy.

Table 1 Execution time of different methods

S. No.	Methods	Execution time (s)
1	IPO	0.7
2	IPNN	3.5
3	IPSMC	4.7

Table 2 Computational complexity in %

S. No.	Methods	Computational complexity in %
1	IPO	22
2	IPNN	62
3	IPSMC	73

Table 3 Computational complexity in %

S. No.	Methods	Computational cost (%)
1	IPO	23
2	IPNN	68
3	IPSMC	81

Table 4 Computational cost of perturbed data of different fashion

Computational cost of perturbed data (using noise y)					
	Original data	Perturbed data using uniform distribution on (0, 1)	Perturbed data using uniform distribution on (− 1, 1)	Difference between perturbed data based on distribution (0, 1) and original data	Difference between perturbed data based on distribution (− 1, 1) and original data
NB	0.13	0.08	0.06	− 0.05	− 0.07
MLP	34.58	38.59	36.67	4.01	2.09
CART	5.06	13.45	17.61	8.39	12.55

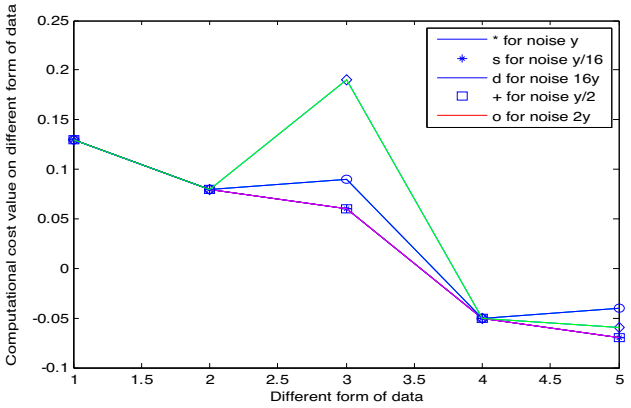

Fig. 1 Computational cost value generated by different form of data using NB classifier

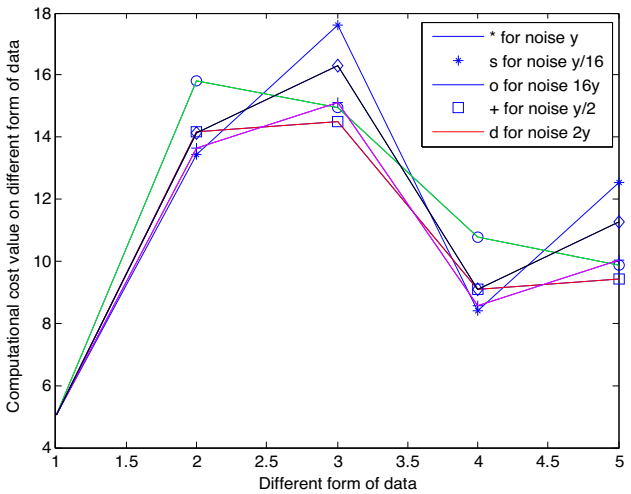

Fig. 2 Computational cost value generated by different form of data using CART classifier

If we multiply every natural number to a noise distribution, it will improve privacy by two, such as $x \in [x - 2y, x + 2y]$ and length of privacy would be 4y and so on. Similarly, we get $x \in [x - ny, x + ny]$ with privacy length 2ny. When $n \to \infty$, then $2ny \to \infty$, the privacy length is very large. Under this circumstance it is difficult to find the original data. Thus, it needs certain limit of range of the

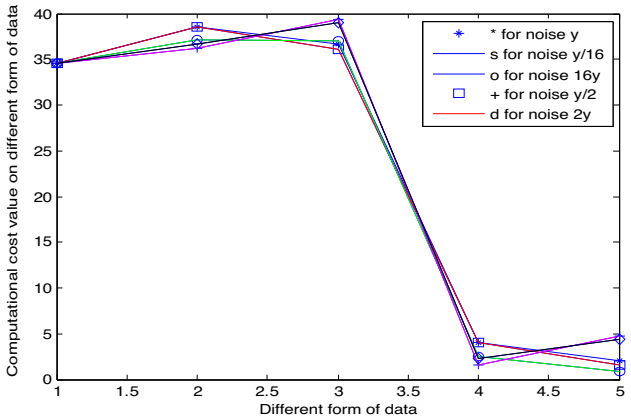

Fig. 3 Computational cost value generated by different form of data using MLP classifier

interval, otherwise it would be worthless. This certain range of interval can be chosen by user according to desired privacy. There is a set of privacy that the data owner can choose using this parameter. When noise grows, privacy increases and data robustness will be stronger. As per data owner requirement, various types of perturbed data are used for computing cost evaluation as in Table 4.

In response to data owner demands, the tests are carried out. Simplification of the privacy issue is used for the optimization of the owner's required privacy level. Both privacy and cost are taken into consideration when looking at the proposed model. Each piece of optimization component is extracted from the following: (a) Privacy: The privacy is required for data owner is shown (Figs. 4, 5) and determines the type of information. Privacy is gauged by (interval length/original data). Different computational results are created as per confidential interval as shown in Table 5 which presents the output through $C_{HP} * U_{HP}$. (b) Cost: We also evaluated the cost U_{CC} on low computational cost (i.e., as a high priority). For the uniform distributions between values (− 1, 1) and (0, 1), we found the evaluation results as shown in Fig. 6. It is estimated that perturbed data is evaluated to analyze as in Table 4. The C_{CC} measurement method is considered in certain set of data, but sufficiently accurate for our purposes. The

importance of computational cost for classifiers centered on changed data is accepted as per our experiments. We also considered the measurements of various computational cost as per membership function as Eq. (19):

$$\mu_x = \begin{cases} 0.2 & if\ HCC \\ 0.7 & if\ MCC \\ 1.0 & if\ LCC \end{cases} \tag{19}$$

where the MCC stands for a moderate computational cost, whereas the HCC stands a higher significant computation cost, and the LCC stands a lower computation cost. The analytical cost efficiency factor considers both original data and perturbed data as per priority basis.

We used two parameters such as d_k and $(1 - \alpha)$ for evaluating the range of both privacy and computational cost as Eq. (18). Experiments conducted utilizing different conditions where data owners' privacy is varied as per demand. After all this information has been obtained, the optimization process is run. The constraints are needed in the estimation of optimization problem. The demands of the data owner are not necessarily satisfied by simple α-cuts value. At this stage, the α-values (0.2, 0.4, 0.6, 0.8, and 1.0) are considered for optimization for various degrees of privacy. The α-cuts depends on the amount of data required by the data owners. Computational costs can differ in proportion to privacy costs. Finding the correct solution

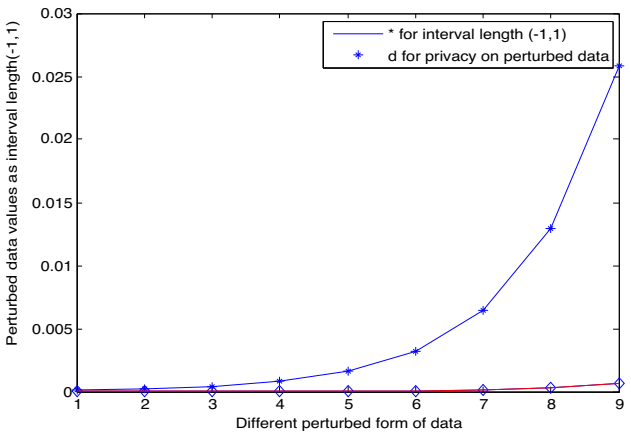

Fig. 4 Interval length (− 1, 1) and its privacy

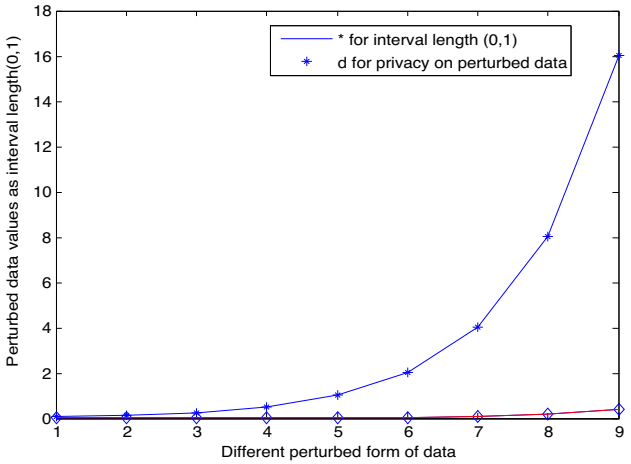

Fig. 5 Interval length (0, 1) and its privacy

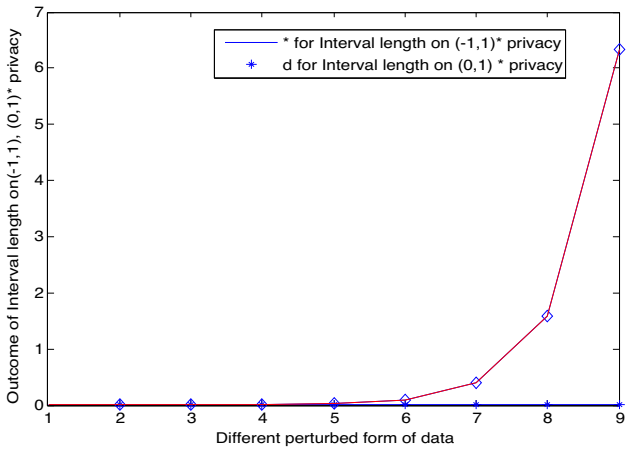

Fig. 6 Results of Interval length (− 1, 1) and (0, 1) with its privacy

Table 5 Interval length * privacy = C_{HP} * U_{HP}

	y/16	y/8	y/4	y/2	y	2y	4y	8y	16y
Interval length on (− 1, 1)* privacy	2.52813E − 10	1.011E − 09	4.04E − 09	1.62E − 08	6.46E − 08	2.59E − 07	1.03E − 06	4.14E − 06	1.65E − 05
Interval length on (0, 1) * privacy	9.673E − 05	0.000387	0.001547	0.006189	0.024754	0.099018	0.39607	1.58428	6.337122

Table 6 Priority of computational cost

Classifier	Distribution data on (0, 1)	Distribution data on (− 1, 1)
For noise y/16 of distributions between (0, 1) and between (− 1, 1)		
MLP	0.0259134	0.0276778
CART	0.0705716	0.0690608
NB	12.5	16.666667
For noise y/2 of distributions between (0, 1) and between (− 1, 1)		
MLP	0.0276319	0.0254388
CART	0.0733138	0.0661813
NB	12.5	16.666667
For noise y of distributions between (0, 1) and between (− 1, 1)		
MLP	0.0259134	0.0272702
CART	0.0743494	0.0567859
NB	12.5	16.666667
For noise 2y of distributions between (0, 1) and between (− 1, 1)		
MLP	0.02722331	0.0256739
CART	0.0707214	0.0613121
NB	12.5	11.111111
For noise 16y of distributions between (0, 1) and between (− 1, 1)		
MLP	0.0269469	0.0270343
CART	0.0631712	0.0669344
NB	12.5	5.2631579

Table 7 Difference of computational cost among three classifiers

	NB	MLP	CART
For different noise distributions between (0,1)			
y/16	− 0.5	4.01	9.11
y/2	− 0.5	1.61	8.58
y	− 0.5	4.01	8.39
2y	− 0.5	2.23	9.08
16y	− 0.5	2.53	10.77
For different noise distributions between (− 1,1)			
y/16	− 0.07	1.55	9.42
y/2	− 0.07	4.73	10.05
y	− 0.07	2.09	12.55
2y	− 0.04	4.37	11.25
16y	0.6	0.91	9.86

Table 8 α-cut value for noise data

α-cut value	0.2	0.4	0.6	0.8	1
Different noise data	y/16	y/2	y	2y	16y

means that the data owner must choose privacy. So that computational cost is fair and is distributed among data owners. Thus, the data miner sacrifices computing efficiency to prevent an excessive level of intrusion into the data owners' privacy. Computational cost burdens correspond to privacy, i.e., attention carries higher/lower priority than cost i.e., lower cost is considered for higher priorities and high cost is considered lower priorities respectively. The cost is considered the primary objective in the formulation of the optimization problem as shown Table 6.

The computational cost gap between the initial and perturbed classifiers is shown in Table 7. Classifier CART takes more time to evaluate because it's more creative in nature in regression tree. So, the utility of the CART classifier is rather abysmal.

The α-cut values provided in Table 8 represent estimated noise in the various experiments. Noise details are taken into account in all distributions. For the optimality test, utility variables have been regarded as data sets. To calculate one of usefulness, it must add in the computing cost, which is designated as $\{C_{CC} = 0.2$ for high cost, 0.7 for medium cost, and 1.0 for low cost. Priority C_{CC} is estimated as follows: Strong = 1.0, middle = 0.7, and low = 0.2. In this scale, the weighting factors are valued as $a_i^{HP} = \{0.3, 0.5, 0.7\}$ and $a_i^{CC} = \{0.7, 0.5, 0.3\}$ respectively where $a_i^{HP} + a_i^{CC} = 1$. Many of our experiments use the measured weight factors as prototype. However, there could be various weight factors for considerations which could be taken into account to adjust the tests' weights. Using two distinct α-cut values, the test of the experiment indicates that uniform distributions produce acceptable results. Alternatively, a definitive test can be stated as the following. Let $\alpha = 0.2$, $a_i^{HP} = 0.3$, $a_i^{CC} = 0.7$, $C_{HP-} = 0.000101125$, $U_{HP} = 2.50E\text{-}06$, $C_{CC} = 0.3$, $U_{CC-} = 0.0276778$ and perturbed data based on uniform distribution between (− 1, 1), then optimal value of $U_i(X)$ is 0.0058.

In order to serve the data owner's various demands, we have carried out evaluation focused on α-cutting measurements by data holders. Optimal solutions have been found for three sets of expected costs of computing. To account for all the computations, we have used two uniform distributions. For different values of α, several optimal values have been developed. The results of the optimization model that the ideal (optimal) solution occurs when alpha = 1.0. When this distribution is implemented, the data owner gets the highest level of privacy for the minimum amount of computational cost in all scenarios. The efficiency has been maximized when utility function is maximum for uniform distributions between (0, 1). The interval length increases (the difference between the upper and lower bounds is small), the value of the utility function shrinks. Hence it is inversely proportional to each other.

Specifically, the optimum solution happens at a value of α of 1.0 and overall other values of α. i.e., 16y of noise data level, the highest level of privacy is maintained. Indirectly it is presumed that increase in noise data for more privacy reduces the extraction of original data from the perturbed one. The optimum solution depends on the various computational complexity. When computational tools are used to satisfy individual privacy and cost requirements, the computational experience modifies the efficiency of both. Therefore, when dealing with a multiple-objective issue dependent on data owner privacy and cost, optimization of above model is the solution.

The uniform distribution is suitable for the evaluation on the (0, 1) and (− 1, 1). It is seen that the optimum value function takes on a different type of uniform distribution (0, 1) and (− 1, 1). Thus, the data miner will determine which distribution fits his needs by selecting between two options: either (0, 1) and (− 1, 1). Interesting to remember that NB classification still yields better results as in Table 6. Thus, NB serves as an important tool in analyzing meaning. Additionally, including some kind of weight variables makes it possible to rate the functional/optimal value. The computational cost disparity between original and perturbed data is negligible (it is seen in Table 4). The accuracy of each classifier on all of the various weighting variables, at α is 1.0. For all weight considerations, the NB classifier has superior results in testing. The efficiency of classifiers, however, differs from factor to factor.

On uniform distribution, the best-optimal solution has been found at 1.0. Thus, at α = 1.0, the optimum solution (or functional value of the utility function) is approximately 1.91 1.93, 10.65, 4.44, 4.45, 8.19, 3.17, 3.19, and 9.42. The optimal sets for privacy and computational cost are {0.4} and {12.5, 0.063, 0.027} which satisfy our optimization model.

6.5 Statistical evaluations of propose model

The individual privacy model is performed using different statistical approaches such as (a) mean, (b) variance, (c) standard deviation as per existing methods and classifiers. As per our statistical evaluation, the comparative mean, variance and standard deviations are evaluated through various classifiers as shown in Table 9. We considered three classifiers as (a) NB, (b) MLP, (c) CART, and three methods as (a) IPO, (b) IPNN, (c) IPSMC.

The statistical values as Table 9 determines the closeness of privacy when data owner want to release his data. Although, privacy measured by using confidence interval as Tables 4 and 5 as per perturbed data, but various classifiers are evaluated using statistical measurements for closeness of released original data. Thus, how much loss of data can be estimated as above statistical measurements.

Table 9 Comparative mean, variance and standard deviation among methods and classifiers

Methods	NB	MLP	CART
Mean value			
IPO	96.1	93.1	92.1
IPNN	90.1	82.3	87.2
IPSMC	80.1	73.2	84.3
Variance			
IPO	0.71	0.79	0.72
IPNN	0.86	0.91	0.85
IPSMC	0.78	0.87	0.93
Standard deviation			
IPO	0.51	0.61	0.71
IPNN	0.63	0.71	0.73
IPSMC	0.74	0.88	0.85

To make balance of privacy and loss of information, above comparative methods are helped to maintain good determination among utility of privacy and computational cost.

7 Conclusions

In this paper, a multi-objective optimization-based privacy model in data mining is developed with the performance of privacy and computational cost which are considered as two components of privacy preserving data mining. Above two components are explained as per our proposed model with the help of optimization model and various classifier approaches. In this model, we emphasized on maintaining individual privacy during computational processing when data owners want to maintain their own privacy as per computational cost. Thus, we have maintained privacy of above items in such a way that both components are reciprocal to each other where data owners decide how much privacy need to maintain. Various privacy requirements produce the best response concerning consumer demand. Different classifiers and utility parameters have been employed for our optimization model of privacy. The computational learning test has considered for optimal problem in many different real-time tests using the UCI Repository datasets. Since our goal is to maximize privacy with minimum computational cost, the optimal model offers the appropriate solution based on weightage of parameters. The evaluation performance is well analyzed as per data owner demand. Our model can be used for various field such as selling of product based on price and quality, milage of vehicle as per utility of oil and distance covering etc. We have planned to make healthcare system using machine learning approaches based on our model in future.

Author contributions All authors have equal contribution.

Funding No funding.

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest No conflict of interest.

Ethical approval None.

References

1. Alzubi, O.A., Alzubi, J.A., Al-Zoubi, A.M., Hassonah, M.A., Kose, U.: An efficient malware detection approach with feature weighting based on Harris Hawks optimization. Clust. Comput. J. (2021). https://doi.org/10.1007/s10586-021-03459-1
2. Movassagh, A.A., Alzubi, J.A., Gheisari, M., Rahimi, M., Mohan, S., Abbasi, A.A., Nabipour, N.: Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model. J. Ambient Intell. Humniz. Comput. (2021). https://doi.org/10.1007/s12652-020-02623-6
3. Gheisari, M., Alzubi, J., Zhang, X., Kose, U., Saucedo, J.A.M.: A new algorithm for optimization of quality of service in peer to peer wireless mesh networks. Wirel. Netw. **26**, 4965–4973 (2020). https://doi.org/10.1007/s11276-019-01982-z
4. Cadenas, J.M., Verdegy, J.L.: A primer on fuzzy optimization models and methods. Iran. J. Fuzzy Syst. **3**(1), 1–21 (2006)
5. Herrera, F., Verdegay, J.L.: Three models of fuzzy integer linear programmining. Eur. J. Oper. Res. **83**, 581–593 (1995)
6. Bayardo, R. J., Agrawal, R.: Data privacy through optimal k-anonymization. In: Proceedings of ICDE'05, Washington, DC, USA, IEEE Computer Society, pp. 217–228, 2005
7. Gheisari, M., Najafabadi, H.E., Alzubi, J.A., Gao, J., Wang, G., Abbasi, A.A., Castiglione, A.: OBPP: an ontology-based framework for privacy-preserving in IoT-based smart city. Future Gener. Comput. Syst. **123**, 1–13 (2021). https://doi.org/10.1016/j.future.2021.01.028
8. Alzubi, O.A., Alzubi, J.A., Shankar, K., Gupta, D.: Blockchain and artificial intelligence enabled privacy preserving medical data transmission in internet of things. Trans. Emerg. Telecommun. Technol. **32**(2), 1–14 (2021)
9. Krivitski, D., Schuster, A., Wolff, R.: A local facility location algorithm for large-scale distributed systems. J. Grid Comput. **5**(4), 361–378 (2007)
10. Xiao, Y., Xiong, L., Fan, L., Goryczka, S., Li, H.: DPCube: differentially private histogram release through multidimensional partitioning. Trans. Data Priv. **7**(3), 195–222 (2014)
11. Clifton, C., Tassa, T.: On syntactic anonymity and differential privacy. Trans. Data Priv. **6**(2), 161–183 (2014)
12. Song, J., Wang, W., Gadekallu, T.R., Cao, J., Liu, Y.: EPPDA: an efficient privacy-preserving data aggregation federated learning scheme. IEEE Trans. Netw. Sci. Eng. (2022). https://doi.org/10.1109/TNSE.2022.3153519
13. Wenjie, D., Yang, W., Zhou, J., Shi, L., Chen, G.: Privacy preserving via secure summation in distributed Kalman filtering. IEEE Trans. Control Netw. Syst. (2022). https://doi.org/10.1109/TCNS.2022.3155109
14. Bhuyan, H.K., Kamila, N.K., Pani, S.K.: Individual privacy in data mining using fuzzy optimization. Eng. Optim. (2021). https://doi.org/10.1080/0305215X.2021.1922897
15. Perez, I.J., Alonso, S., Cabrerizo, F.J., Lu, J., Herrera-Viedma, E.: Modelling Heterogeneity among Experts in Multi-criteria Group Decision Making Problems, pp. 55–66. Springer-Verlag, Berlin (2011)
16. Dutta, D., Murthy, S.: Multi-choice goal programming approach for a fuzzy transportation problem. IJRRAS **2**(2), 132 (2010)
17. Jimenez, F., Cadenas, J.M., Sanchez, G., Gomez-Skarmeta, A.F., Verdegay, J.L.: Multiobjective evolutionary computation and fuzzy optimization. Int. J. Approx. Reason. **43**, 59–75 (2006)
18. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
19. Deb, K.: Multiobjective Optimization Using Evolutionary Algorithms. Wiley, Hoboken (2001)
20. Tanaka, H., Okuda, T., Asai, K.: On fuzzy mathematical programming. J. Cybern. **3**(4), 37–46 (1974)
21. Mukherjee, S., Chen, Z., Gangopadhyay, A.: A fuzzy programming approach for data reduction and privacy in distance based mining. Int. J. Inf. Comput. Secur. **2**(1), 27–47 (2008)
22. Asuncion, A., Newman, D.: UCI machine learning repository, 2007.
23. Bhuyan, H.K., Kamila, N.K.: Privacy preserving sub-feature selection based on fuzzy probabilities. Clust. Comput. **17**(4), 1383–1399 (2014)
24. Bhuyan, H.K., Mohanty, M., Das, S.R.: Privacy preserving for feature selection in data mining using centralized network. Int. J. Comput. Sci. Issues (IJCSI) **9**(3), 434–440 (2012)
25. Teo, S.G., Cao, J., Lee, V.C.S.: DAG: a general model for privacy-preserving data mining. IEEE Trans. Knowl. Data Eng. **32**(1), 40–53 (2020)
26. Kim, S., Shin, H., Baek, C., Kim, S., Shin, J.: Learning New Words from Keystroke Data with Local Differential Privacy. IEEE Trans. Knowl. Data Eng. **32**(3), 479–491 (2020)
27. Christen, P., Ranbaduge, T., Vatsalan, D., Schnell, R.: Precise and fast cryptanalysis for bloom filter based privacy-preserving record linkage. IEEE Trans. Knowl. Data Eng. **31**(11), 2164–2177 (2019)
28. Bhuyan, H.K., Dash, S.K., Roy, S., Swain, D.K.: Privacy preservation with penalty in decentralized network using multiparty computation. Int. J. Adv. Comput. Technol. (IJACT) **4**(1), 297–303 (2012)
29. Bhuyan, H.K., Kamila, N.K., Dash, S.K.: An approach for privacy preservation of distributed data in peer to-peer network using multiparty computation. Int. J. Comput. Sci. Issues (IJCSI) **3**(8), 424–429 (2011)
30. Kamila, N.K., Jena, L.D., Bhuyan, H.K.: Pareto-based multiobjective optimization for classification in data mining. Clust. Comput. (Springer) **19**, 1723–1745 (2016)
31. Bhuyan, H. K., Madhusudan Reddy, C. V.: Sub-feature selection for novel classification. In: IEEE Explore. April, 2018.
32. Bhuyan, H.K., Ravi, V.K.: Analysis of sub-feature for classification in data mining. IEEE Trans. Eng. Manage. (2021). https://doi.org/10.1109/TEM.2021.3098463
33. Bhuyan, H. K., Raghu Kumar, L., Reddy, K. R.: Optimization model for Sub-feature selection in data mining. In: IEEE Explore. (2020)
34. Alazab, M., Lakshmanna, K., Reddy, T., Pham, Q.V., Maddikunta, P.K.: Multi objective cluster head selection using fitness averaged rider optimization algorithm for IoT networks in smart cities. Sustain. Energy Technol. Assess. **43**, 1–19 (2021)
35. Kumar, R., Kumar, P., Tripathi, R., Gupta, G.P., Gadekallu, T.R., Srivastava, G.: Sp2f: a secured privacy-preserving framework for smart agricultural unmanned aerial vehicles. Comput. Netw. **187**, 1–15 (2021)
36. Kumar, P., Kumar, R., Srivastava, G., Gupta, G.P., Tripathi, R., Gadekallu, T.R., Xiong, N.N.: PPSF: a privacy-preserving and secure framework using blockchain-based machine-learning for

IoT-driven smart cities. IEEE Trans. Netw. Sci. Eng. **8**(3), 2326–2341 (2021)

Hemanta Kumar Bhuyan received PhD degree in Computer Science and Engineering from Sikshya 'O' Anusandhan (SOA) University, Odisha, India and M.Tech. from Utkal University, Odisha, India in 2016 and 2005 respectively. He is currently an Associate Professor in Vignan's Foundation for Science, Technology & Research (Deemed to be University), Guntur, Andhra Pradesh, India. He has published many research papers in national/international journals of repute. His research interests include distributed data mining, privacy preservation, feature selection and fuzzy system.

Vinayakumar Ravi is an Assistant Research Professor at Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia. My previous position was a Postdoctoral research fellow in developing and implementing novel computational and machine learning algorithms and applications for big data integration and data mining with Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA from September, 2019 to September, 2020. He received the Ph.D. degree in computer science from Computational Engineering & Networking, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India. His Ph.D. work centers on Application of Machine learning (sometimes Deep learning) for Cyber Security and discusses the importance of Natural language processing, Image processing and Big data analytics for Cyber Security. His current research interests include applications of data mining, Artificial Intelligence, machine learning (including deep learning) for biomedical informatics, Cyber Security, image processing, and natural language processing. More details available at https://vinayakumarr.github.io/. He has more than 50 research publications in reputed IEEE conferences, IEEE Transactions and Journals. His publications include prestigious conferences in the area of Cyber Security, like IEEE S&P and IEEE Infocom. He has given many invited talks on deep learning applications in IEEE conferences and Industry workshops in 2018. He has got a full scholarship to attend Machine Learning Summer School (MLSS) 2019, London. Dr. Ravi has served as a Technical Program Committee (TPC) member at international conferences including SSCC Symposium, IEEE TrustCom-2020, and IEEE SmartData-2020. He is an editorial board member for Journal of the Institute of Electronics and Computer (JIEC), International Journal of Digital Crime and Forensics (IJDCF), and he has organized a shared task on detecting malicious domain names (DMD 2018) as part of SSCC'18 and ICACCI'18. He received the Chancellor's Research Excellence Award in AIRA 2021 and was included in the World's Top 2% Scientists by Stanford University published in PLoS Biology.

M. Srikanth Yadav is an assistant professor in Vignan's Foundation for Science, Technology & Research. His areas of research are machine learning, data mining, security and privacy.