Methods for Responsible Machine Learning:

Privacy-Preserving Data Mining in Banking Applications

COMP 551 Final Project

Raylene MacDonald
Email: raylene.macdonald@mail.mcgill.ca
Student ID: 260557908

Amen Memmi
Email: amen.memmi@mail.mcgill.ca
Student ID: 260755070

Hui Lee Ooi
Email: hui.ooi@mail.mcgill.ca
Student ID: 260757354

***Abstract*—This project presents various methods of privacy preserving data mining in a banking application. We investigate the effects of these algorithms on the binary classification task of identifying whether credit card clients will default on a payment. Several of the methods examined are found to preserve patterns in the unmodified data, resulting in a minimal decrease in classification accuracy.**

I. Introduction

In recent years, data mining has become increasingly employed across a variety of disciplines, including the medical field, the banking industry, and in security intelligence. Many applications in such fields involve sensitive personal data being collected by companies and governments; this poses a security risk in the case of a data breach by malicious groups. There has also been concern about misuse of personal data by private companies. To combat these issues, the practice of Privacy-Preserving Data Mining (PPDM) has emerged, which involves several different approaches including cryptographic methods for preventing security breaches. Secure Multi-Party Computation, which involves distributing data across nodes which independently perform mining on non-identifying partitions of the data, and data perturbation, which is the focus of this paper.

Data perturbation modifies the original data such that individual records cannot be reconstructed, thus preserving the privacy of persons whose information exists in the database. The intention behind such methods is that sensitive data is collected by a trusted third-party, who applies the modifications and provides the perturbed data to the agency performing data mining. Of course, the utility of the modified data set must be preserved; the objective of privacy-preserving methods is therefore to alter the data to a sufficient extent without significantly impacting the accuracy of the predictive models.

Techniques investigated in this paper are Singular Value Decomposition (SVD), a modified version of it (SSVD), Random Rotation Perturbation, adding matrices of random noise to the feature matrix, as well as K-Anonymity. Several metrics measuring the extent to which the data has been altered are used for comparison of the methods. Additionally, the classification accuracy under various applications of the above-mentioned algorithms is presented to demonstrate how well patterns in the original data have been preserved.

II. Related work

A. Rotation Perturbation

Rotation perturbation is categorized a form of geometric perturbation, where the combination of rotation, translation and addition of random noise is applied to the data for preserving the data quality without jeopardizing data privacy. In the work by [1], it is demonstrated that geometric rotation is used in several classification models. [1] proposed the use of a multi-column privacy model with multi-dimensional perturbation for evaluation of privacy quality. [2] further extended the work by presenting a unified privacy model that aims to tackle three possible attacks, namely direct estimation, approximate reconstruction and distribution-based inference attacks. Meanwhile, [3] presented vertical partitioning of data matrix vertically with application of rotational matrix to perturb each sub-set matrix.

B. Singular Value Decomposition

Presented in [4] are several methods for privacy preservation, including Singular Value Decomposition (SVD), and the extended Sparsified Singular Value Decomposition (SSVD). SVD is most commonly used for dimension reduction; the parameter k is chosen to be less than the number of features, and allows the algorithm to construct k artificial attributes which can be said to approximate the original feature set. Also introduced in [4] is the method of adding matrices of random noise to the data matrix. The values of the noise matrix are distributed according to a probability distribution; in this case, Normal and Uniform distributions are investigated. The four methods described here are implemented in this paper.

A more secure method of privacy preservation using SVD is presented in [5], using data partitioning, which allows for distributed processing of the sensitive data. Both vertical and horizontal partitions are considered. The authors describe algorithms allowing SVD to be performed securely on the partitioned data, including Secure Matrix Multiplication [5] and a privacy-preserving QR algorithm.

C. Banking Application: Default Classification

A particular application of data mining is determining which clients of a bank should be given credit. This is done in [6] by examining the accuracy of various classifiers in predicting the default status of credit card clients. The classification methods implemented are K-Nearest Neighbor, Logistic Regression,

Discriminant Analysis, Naive Bayes, Neural Networks and Decision Trees. Among those, Neural Networks are found to have the highest accuracy. A new technique for estimating the real probability of default is presented: "Sorting Smoothing Method" [6], which uses linear regression to compare the results of the various classifiers. We use their dataset, from [7], in this paper.

III. Problem representation

A. Features

The dataset used in this paper is taken from UCI's Machine Learning Repository [7]. We attempt to classify the default status of 30,000 credit card clients in Taiwan by examining various personal attributes as well as recent payment history. An outline of the features used is presented in Table I. The target variable is binary, with 1 representing a default on the next month's bill and 0 indicating a timely payment. These types of predictions are valuable for banks attempting to determine which clients to grant credit to, and how much they should be given.

B. Data Normalization

Standard score normalization from Equation 1 is performed on data to scale the values [0,1]. This normalization step is performed on continuous features ($X1$, $X5$ and $X12$ to $X23$).

$$X_{norm} = \frac{x - \overline{x}}{\sigma}. \tag{1}$$

where $\overline{x}$ is the mean of the feature and σ is the standard deviation of the feature. X_{norm} is the normalized features.

C. Imbalanced Classes

The distribution of classes in the data is imbalanced, with 23364 samples belonging to class "0" and 6636 samples belonging to class "1". This results in a baseline accuracy of approximately 78%, achieved by a classifier predicting only class 0. Since the best accuracy we were able to obtain was 81%, we wanted a larger window to analyze the effect of privacy-preserving methods under various parameter settings. We therefore decided to balance the classes, such that our baseline would be 50%.

For handling the imbalance of class in the data, the data points from both classes are sampled to obtain a new set of data with balanced class. Samples from class "0" will be under-sampled (sample without replacement) whereas samples from class "1" will be over-sampled (sample with replacement) as shown in Algorithm 1. The implementation of Rotation Perturbation approach is performed using the sampled data.

Algorithm 1 Sampling algorithm

```
procedure SAMPLING
loop for each class label:
    if count(class) ≥ n_samp then
        Sample without replacement
    else
        Sample with replacement
    for current sample data count < n_samp do
        append sampleData
```

IV. Methods

A. Rotation Perturbation

The work on Rotational Perturbation implemented in this project is inspired from the proposed approach by [1]. Rotation matrix is defined as a matrix that perform rotation in Euclidean space, as defined by Equation 2. Through an angle θ about the origin of the Cartesian coordinate system, the matrix rotates points in the xy-Cartesian plane.

$$R(\theta) = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{2}$$

where θ is the rotational angle in the perturbation.

Rotation matrix should satisfy the following property in Equation 3:

$$R * R^T = R^T * R = I \tag{3}$$

where I is the identity matrix and R^T denotes the transpose of R.

X_{mxn} denotes the data consisting of n data (records) and d columns (attributes) with $X = [x_1, x_2, ...x_n]$, where x_i is data tuple belonging to a class.

With data X_{mxn} of dimension n attributes, the rotation matrix $Rmatrix_{nxn}$ is required with the dimension nxn. In this project, $Rmat_{nxn}$ is obtained by repetitive concatenation of the rotation matrix in Equation 2. In the case where n is a odd number, the final row or column of the matrix is joined with randomly drawn row or column from matrix in Equation 2. Perturbed data $X_{rotation}$ is generated from the multiplication of X with $Rmat_{nxn}$ as shown in Equation 4. Privacy is preserved since the values in the matrix X_{rot} would be different compared to X.

$$X_{rot} = Rmat_{nxn} * X \tag{4}$$

B. K-Anonymity

1) Definition: K-anonymity is a privacy preserving technique that was first introduced by Latanya Sweeney in a paper published in 1998 [8] as an attempt to solve the problem: "Given person-specific field-structured data, produce a release of the data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful." A release of data is said to have the k-Anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release. This technique consists in applying series of transformations on the attributes until each row (data instance, client) is identical with at least k-1 other rows. Transformations include suppression and generalization: Suppression means removing downright the quasi-identifying individual attributes by replacing them

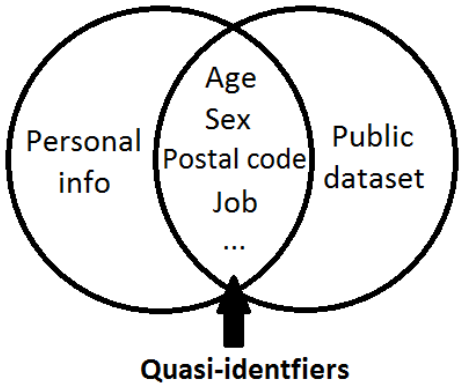
TABLE I: Dataset Attributes

Attribute	Description
X1: Credit	Maximum credit given, in USD
X2: Gender	1 = female, 2 = male
X3: Education	1 = graduate school, 2 = university 3 = high school, 4 = others
X4: Marital Status	1 = married, 2 = single, 3 = others
X5: Age	In years
X6 - X11: History of past payment	For months of April to September, 2005 -1 = pay duly 1 = payment delay for one month 2 = payment delay for two months . . . 9 = payment delay for nine months or more
X12 - X17: Amount of bill statement	For months of April to September, 2005 In USD
X18 - X23: Amount of previous payment	For months of April to September, 2005 In USD

with stars " * ". Generalization on the other hand replaces individual attributes with a broader category according to a predefined hierarchy: Say generalizing the age from 26 to a decade interval of $[20-30]$; or generalizing occupations like *Police officer* and *Administrative employee* to *Governmental post*. Suppression could be wxpressed as an additional final level of generalization, "*" level.

K-Anonymity is able to prevent definite database linkages. At worst, the data released narrows down an individual entry to a group of k individuals. Unlike Output Perturbation models, K-Anonymity guarantees that the data released is accurate.

2) Quasi-identifiers: When de-identifying records, many people assume that removing names and addresses (direct identifiers) is sufficient to protect the privacy of the persons whose data is being released. The problem of de-identification involves those personal details that are not obviously identifying. These personal details, known as quasi-identifiers, include the persons age, sex, postal code, profession, ethnic origin and income (to name a few).



In our case, we are considering all attributes as quasi-identifiers and thus we are applying K-Anonymization on the whole set of attributes.

3) Implementation using ARX anonymization tool: ARX is a comprehensive open source software for anonymizing sensitive personal data. It allows to apply few anonymization techniques using a wide variety of parameters and enables us to evaluate performances and select the optimal solution (set of transformations) that minimizes the defined utility measure. After importing our data set to the software, we defined a generalization hierarchy for each attribute. For example, generalizing ages to 5 years intervals in the first level, decades on the second and so on as it can be seen in figure 1. Another example is generalizing attributes of past history of payment, represented by integers $\in \{-1, 0, 1, ..., 9\}$, to sets of three successive values. Details of the hierarchy trees could be consulted on the ARX project file attached. After that we specify the privacy model to be implemented, 5-Anonymity in our case and then we define the utility measure we are optimizing. In this stage, the software scans all the solution space of all eventual combinations of suppression and generalization to different levels and then selects the one maximizing our utility measure.

4) Performance metrics: Our aim is to evaluate K-Anonymity performances on two different levels:

• In one hand, we want to analyze results of anonymization process in terms of privacy gain for the dataset, i.e, how anonymous and safe are the records of clients against identity disclosure. Re-identification risks estimates are provided for three different attacker models: (1) the prosecutor scenario, (2) the journalist scenario and (3) the marketer scenario. In the prosecutor model it is assumed that the attacker already knows that data about the targeted individual is contained in the data set. In the journalist model, such background knowledge is not assumed. In the marketer model, it is assumed that the attacker is not interested in re-identifying a specific individual but that she aims at attacking a larger number of individuals. ARX provides results for those three models in addition to few other metrics.

• On the other hand, to investigate the effect of anonymizing the data using k-Anonymity on the classification accuracy, we

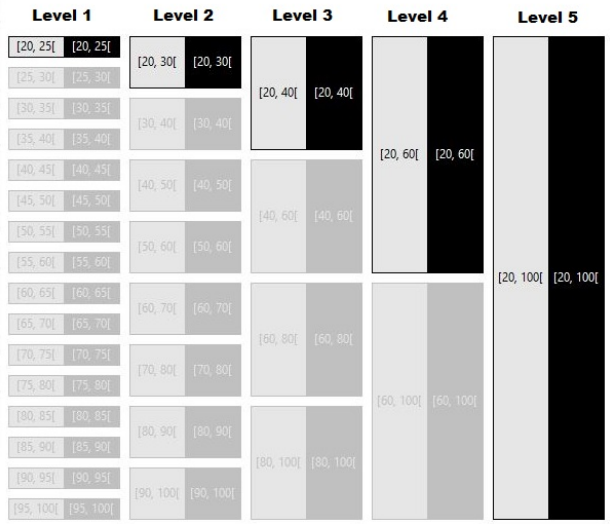


Fig. 1: Age hierarchy for K-anonymity

implemented 3 classifiers: Logistic regression, Naive Bayes as well as Random Forest classifiers. We compare then classification performances before and after K-Anonymization for different values of K.
Results of the anonymization process will be discussed in the following section.

C. Addition of Random Noise

Following methods implemented in [4], the original feature matrix, A, was randomized by adding to it matrices of probabilistically distributed noise, N, to obtain the matrix $\overline{A}$. The intent is to alter the entries of each record enough such that individual samples become unrecognizable. Clearly the original A is not reconstructible, as the values of N have been drawn from a random distribution. We define the perturbed matrices $\overline{A_u}$ and $\overline{A_n}$ as follows:

$$\overline{A_u} = A + N_u \tag{5}$$

$$\overline{A_n} = A + N_n \tag{6}$$

where the entries of N_u are uniformly distributed in the $[0, 1]$ range, and the entries of N_n have been drawn from a Normal distribution with parameters mean μ and standard deviation σ.

D. Singular Value Decomposition (SVD)

SVD is a matrix factorization technique commonly used in data mining. It is generally used for dimension reduction or Principal Component Analysis, to extract the most valuable information from a large feature space. Here it is used as a method of data distortion for privacy preservation. The singular value decomposition of the $n\ x\ m$ matrix A is [9]

$$A = U\Sigma V^T \tag{7}$$

where U is an $n\ x\ n$ orthonormal matrix, Σ is an $n\ x\ m$ diagonal matrix and V^T is an $m\ x\ m$ orthonormal matrix. The number of nonzero entries in Σ is equal to the rank of A, and are arranged in descending order. As such, the highest entries of Σ correspond to the maximal variation among features in A. We define the matrix A_k as follows [9]:

$$A_k = U_k \Sigma_k V_k^T \tag{8}$$

where U_k has the first k columns of U, Σ_k has the first k nonzero diagonals of Σ, and V_k^T has the first k rows of V^T. The dimension of our new dataset A_k has been reduced, but can be said to reasonably approximate the original matrix A. Lower values of k provide increased privacy, at a cost to classification accuracy.

E. Sparsified Singular Value Decomposition (SSVD)

The authors of [9] have proposed an alternative to simple SVD, wherein the matrices U_k and V_k^T are sparsified. Values in these matrices whose magnitude is smaller than a threshold ϵ are set to zero, in what is referred to as the dropping operation. We denote $\overline{U_k}$ as the matrix U_k with entries $|u_{ij}| < \epsilon$ dropped, and similarly for $\overline{V_k^T}$. We can then define the twice-distorted data matrix $\overline{A_k}$ as

$$\overline{A_k} = \overline{U_k}\Sigma_k\overline{V_k^T}. \tag{9}$$

F. Privacy Metric: Value Difference

Several methods for measuring the degree of privacy in a distorted dataset are presented in [4]; here we introduce the simplest, Value Difference (VD). This metric is used to quantify by how much a privacy-preserving algorithm has altered the original values of the data. We use the VD in conjunction with the preserved utility, or accuracy, to compare the various privacy methods. The Value Difference is defined as follows:

$$VD = \frac{\|A - \overline{A}\|}{\|A\|} \tag{10}$$

where A is the original data, $\overline{A}$ is the distorted data, and the norm used is the Frobenius norm.

V. Testing and validation

A. Data Pre-Processing

With the normalized data, K-nearest neighbor (KNN) with $(k = 1)$ classification is performed on the un-sampled and sampled data, as shown in Table II.

TABLE II: Comparison of classification performance before and after sampling to handle imbalanced data

Data	Accuracy	Precision	Recall
Unsampled	0.729	0.388	0.395
Sampled	0.837	0.790	0.917

Different values of k ranging from 1 to 21 are tested for KNN classification on sampled data as shown in Table III. Only odd numbers are considered for k parameter since it is a binary classification task. It is observed from Table III

that with the exception of $k = 1$ (highest accuracy), the classification accuracy varies between 0.711 to 0.758. Hence, $k = 1$ is chosen for all further tests invoiving KNN.

TABLE III: Comparison of k for K nearest neighbor classifier on sampled data

k	Accuracy	Precision	Recall
1	0.837	0.790	0.917
3	0.758	0.722	0.835
5	0.727	0.703	0.782
7	0.720	0.703	0.758
9	0.716	0.705	0.739
11	0.715	0.708	0.731
13	0.710	0.706	0.718
15	0.712	0.710	0.712
17	0.712	0.713	0.707
19	0.710	0.714	0.698
21	0.711	0.717	0.694

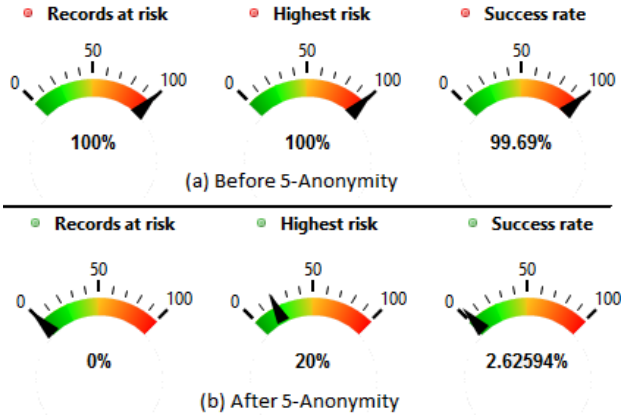


Fig. 2: Prosecutor re-identification risk before and after5-Anonymity

B. Rotation Perturbation Results

KNN classification results with ten fold cross validation on data with rotation perturbation is presented in Table IV with different values of θ ranging from 5 to 50.

Asides from classification performance in terms of accuracy, precision and recall, three privacy metrics: unified column privacy metric $Privacy_{UnitedColumn}$, minimum privacy metric $Privacy_{min}$ and average privacy metric $Privacy_{avg}$ [1], [3]) are also used to evaluate the performance of rotation perturbation in preserving privacy.

$$s_i = \frac{1}{max(\mathbf{X}_i) - min(\mathbf{X}_i)}. \tag{11}$$

$$t_i = \frac{min(\mathbf{X}_i)}{max(\mathbf{X}_i) - min(\mathbf{X}_i)}. \tag{12}$$

where X_i denotes the i_{th} column of data X for Equation 11 and Equation 12.

$$\mathbf{X}_{si} = s_i(\mathbf{X}_i - t_i). \tag{13}$$

$$Privacy_{UnitedColumn} = Var(\mathbf{X}'_{si} - \mathbf{X}_{si}). \tag{14}$$

where $\mathbf{X}'_{si}$ is the perturbed data and $\mathbf{X}_{si}$ is the normalized data computed by Equation 13.

Assuming that all features in the data contribute equally to the classification, $Privacy_{avg}$ and $Privacy_{min}$ are defined in Equation 15 and 16 respectively.

$$Privacy_{avg} = mean(\mathbf{X}'_{si} - \mathbf{X}_{si}). \tag{15}$$

$$Privacy_{min} = min(\mathbf{X}'_{si} - \mathbf{X}_{si}). \tag{16}$$

From the results tabulated in Table IV, it is observed that the introduction of Rotation Perturbation h

C. K-Anonymity

For simulation, unless otherwise stated, we are testing 5-anonymity and loss (in accuracy) as utility measure. Configurating ARX as previously detailed, it performed about 80 billions transformation of the dataset and was able to come out with a solution space of 48 possible solutions among which the optimal transformation: $[7, 1, 2, 4, 0, 0, 0, 0, 1, 1, 4, 0, 0, 0, 0, 0]$, which means that the first attribute has been generalized to the seventh level, the second to the first level, ..., the fifth with no transformation and so on. In the following, we are evaluating privacy and classification performances.

1) Privacy gain: ARX allows various privacy risks analyzed. These include re-identification risks for the prosecutor, journalist and marketer attacker models, which we are considering in our case, as well as other metrics we are not including in our analysis like population uniqueness and HIPAA identifiers. Figure 2 highlights the ability of the privacy technique to reduce the highest prosecutor risk from 100% to 20% with a reduction of the average risk from 99.69% to just 2.62%, meaning that, on average, among the 30,000 clients only 786 may be subject to eventual de-identification. Journalist and Marketer risks' models had similar numbers. Results are summarized in table V.

TABLE V: Re-identification risks before and after applying 5-Anonymity

Measure	[%] Before	[%] After
Lowest prosecutor risk	25	0.147
Records affected by lowest risk	0.026	36.227
Average prosecutor risk	99.614	2.625
Highest prosecutor risk	100	20
Records affected by highest risk	99.416	2.411
Estimated prosecutor risk	100	20
Estimated journalist risk;	100	20
Estimated marketer risk	99.69	2.625

As we can see, the anonymization process allowed to reduce the lowest prosecutor risk to just 0.147% affecting more than

TABLE IV: Comparison of classification performance for data perturbed by rotation matrix

θ (degree)	Unperturbed	5	10	20	30	40	50
Accuracy	0.837	0.797	0.797	0.797	0.797	0.798	0.797
Precision	0.790	0.748	0.749	0.750	0.748	0.750	0.747
Recall	0.917	0.895	0.892	0.892	0.895	0.892	0.895
$Privacy_{UnitedColumn}$	-	0.039	0.039	0.039	0.040	0.040	0.040
$Privacy_{min}$	-	0.010	0.008	0.004	-0.001	0.002	0.002
$Privacy_{avg}$	-	0.120	0.119	0.118	0.116	0.117	0.117

third of the records while reducing the population affected by the highest risk (of 20% compared to the initial 99.41%) to only 2.411% . The estimated risks for prosecutor and journalist models are valued at 20%, which means that the risk of identifying a particular client is cut down by fifth. The marketer risk in the other hand is measured by calculating the probability of matching a record in an equivalence class, not a pre-specified person, trying to re-identify as many individuals as possible. An attack can therefore only be considered successful if a larger portion of the records could be re-identified, which explains why the risk is always lower for this kind of attacks. In our case, it is reduced to just 2.5% using 5-Anonymity, which is below the 5% threshold and then insures dataset safety against this attack.

2) Classification performances after anonymization: As previously detailed, in order to test classification performances before and after applying 5-Anonymity. we implemented 3 common machine learning classifiers. The simulations followed these ideas: The dataset was divided into 2 sets: train and test sets. Stratified sampling (using *StratifiedShuffleSplit from sklearn library*) was used as our target value is unbalanced. In each set, we maintained the ratio of zeros over ones, the same ratio as it is in the full dataset (~ 0.75). Accuracy (ACC), Area-Under-Curve (AUC), Precision (PRE) and Recall (REC) were used as performance metrics. Results shows that those parameters are only slightly affected after anonymizing the data. This can be seen in figure 3 for the case of Random Forest classifier. Others classifiers present similar response and details could be found in table VI for all studied classifiers for values of K going from 2 to 100. For $K = 5$, loss is accuracy was less than 1% and less than 5% for all values of K, which presents a great compromise the gain in privacy previously discussed.

D. Singular Value Decomposition

Using a Random Forest classifier from scikit-learn, we were able to achieve an accuracy of 70.2% on the class-balanced training set using the full feature set of 23 attributes. To obtain the least-variance estimate of the accuracy, 5-fold cross-validation is performed. We use this as our baseline with which to compare the accuracy achieved by an SVD-distorted dataset, for varying values of k. The results are presented in Table VII.

We observe that for $k = 20$, there is actually a slight increase in accuracy compared to the untouched data; this indicates that some redundancy in the dataset has been eliminated. SVD is commonly used to increase validation accuracy in highly redundant datasets.

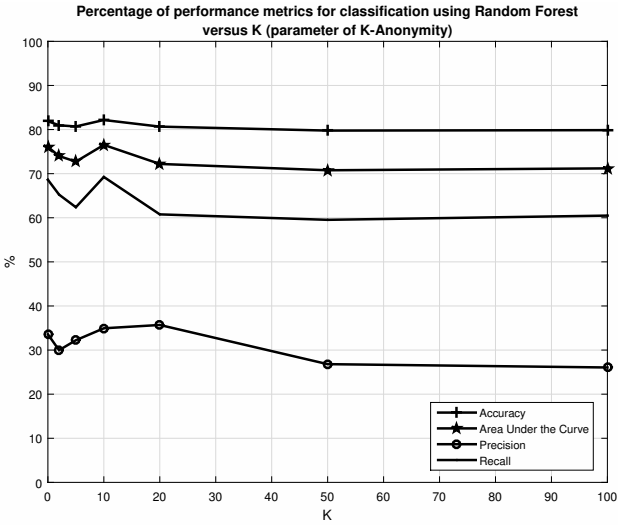


Fig. 3: Performance metrics in percentage for classification using Random Forest versus K parameter of K-Anonymity

E. Sparsified Singular Value Decomposition

The additional step of SSVD is performed for select values of k, as well as on the original data, which is equivalent to $k = 23$. We vary the threshold parameter, ϵ, and compare the accuracies achieved in Table VIII. Based on these results, values for ϵ between 0.01 and 0.1 are reasonable.

F. Addition of Random Noise

The addition of uniformly distributed noise in matrix form yielded accuracy lower than desired, as demonstrated in Table IX. Since the data has been normalized to the [0, 1] range, it is not surprising that noise distributed in the [0, 1] range gives a very low accuracy. However, for the [0, 0.25] range the accuracy is almost equal to the 0.702 achieved on the unperturbed dataset.

Presented below is the accuracy obtained by adding Normally distributed noise, sampled from a distribution with parameters mean μ and standard deviation σ.

As expected, higher values of standard deviation results in lower accuracy, as the data has been perturbed to a higher degree. A mean of 0.5, which is in the center of the range to which the data has been normalized, also gives slightly lowered accuracy.

TABLE VI: Performance metrics for classification using Logistic Regression (LR), Naive Bayes (NB) and Random Forest (RF) versus K parameter of K-Anonymity'

	Accuracy (%)			AUC (%)			Precision (%)			Recall (%)		
K	LR	NB	RF	LR	NB	RF	LR	NB	RF	LR	NB	RF
Original data	80.981	77.123	81.900	61.075	67.240	76.040	69.421	48.391	33.584	25.320	49.496	68.61
2	80.062	76.098	80.996	58.383	54.237	73.995	48.445	39.451	29.969	18.601	15.015	65.245
5	79.773	74.797	80.700	56.230	52.843	72.727	42.887	33.081	32.228	15.663	13.451	62.390
10	78.837	74.467	82.161	57.752	53.780	76.521	32.715	34.239	34.228	17.275	16.662	69.253
20	79.187	76.002	80.666	52.128	55.713	72.204	38.414	41.067	35.692	22.121	19.310	60.769
50	77.721	76.377	79.867	54.144	57.834	70.769	31.125	43.986	26.807	24.771	24.563	59.531
100	74.123	76.461	73.767	54.247	57.761	71.190	30.531	44.214	26.054	24.301	24.207	60.489

TABLE VII: Varying values of k for SVD-distorted data

k	Accuracy
20	0.733
15	0.678
10	0.658
7	0.635
5	0.650
3	0.590
2	0.573
1	0.522

TABLE VIII: Varying values of k, ϵ for SSVD-distorted data

k	ϵ	Accuracy
23	0.001	0.709
23	0.01	0.628
23	0.1	0.587
20	0.001	0.693
20	0.01	0.618
20	0.1	0.484
15	0.001	0.681
15	0.01	0.606
15	0.1	0.361

TABLE IX: Varying ranges for uniformly distributed noise

Range	Accuracy
[0, 1]	0.59
[0, 0.5]	0.62
[0.25]	0.68

TABLE X: Varying parameters of Normal distribution

μ	σ	Accuracy
0	0.1	0.653
0	0.5	0.622
0.5	0.1	0.601
0.5	0.5	0.526

G. Value Difference

For the methods of SVD, SSVD, and random noise, we choose parameter settings which maximize accuracy and compare the Value Difference metric between the original data and the variously perturbed data matrices. Presented in Table XI are only a few combinations of parameter settings for the four methods. An optimization of accuracy vs. value difference could be done to find the optimal method and parameter settings. Additionally, the degree to which the data must be perturbed is domain and problem-specific; the accuracy vs privacy trade-off is a consideration for the designers of a particular classification method to take into account.

TABLE XI: Comparison of Value Differences

Perturbation Method	Accuracy	VD
SVD with $k = 20$	0.733	0.001
SVD with $k = 15$	0.678	0.049
SSVD with $k = 20$, $\epsilon = 0.001$	0.693	1.127
SSVD with $k = 20$, $\epsilon = 0.01$	0.618	1.126
Uniform with range [0, 0.25]	0.681	0.562
Uniform with range [0, 0.5]	0.622	1.122
Normal with $\mu = 0$, $\sigma = 0.1$	0.653	0.390
Normal with $\mu = 0$, $\sigma = 0.5$	0.622	1.949

VI. Discussion

Rotation perturbation allows certain classifiers that are based on geometric properties of data such as KNN to achieve similar accuracy on the transformed data compared to the original data. However, it is observed that the classification results do not differ much upon comparing different angle for rotation perturbation.

K-Anonymity used suppression and generalization of attributes to a higher level of abstraction level to protect quasi identifiers. It yielded great privacy gains since small values of K and this came only with a very slight classification performance reduction. K-Anonymity appears thus as an efficient method for privacy protection. However it requires a knowledge of different attributes proprieties to be able to generate and choose generalization hierarchy. The effect different hierarchy could be investigated to evaluate its effect.

Singular Value Decomposition as a method for privacy-preserving offers a reasonable degree of data perturbation while preserving patterns in the original feature set. Additionally, SVD can result in better classification accuracy on highly redundant datasets, by constructing artificial features which capture the maximal variance among the original attributes. As a further distortion technique, Sparsified SVD yields minimal reduction in accuracy while twice altering the untouched data.

As simpler alternative, adding random noise matrices results in a high Value Difference, indicating a large degree of distortion and thus better privacy preservation. The accuracy obtained by certain combinations of parameter settings is comparable to that of more complex techniques, making it an ideal method when ease of implementation is desired.

VII. Statement of Contributions

We hereby state that all the work presented in this report is that of the authors.

References

[1] K. Chen and L. Liu, "A random rotation perturbation approach to privacy preserving data classification," 2005.

[2] K. Chen and L. Liu, "Privacy preserving data classification with rotation perturbation," in *Data Mining, Fifth IEEE International Conference on*, pp. 4–pp, IEEE, 2005.

[3] Z. Lin, J. Wang, L. Liu, and J. Zhang, "Generalized random rotation perturbation for vertically partitioned data sets," in *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on*, pp. 159–162, IEEE, 2009.

[4] S. Xu and J. Zhang, "Data distortion methods and metrics in a terrorist analysis system," tech. rep., Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security, 2008.

[5] P. S. Y. S. Han, W. Keong Ng, "Privacy-preserving singular value decomposition," tech. rep., IEEE International Conference on Data Engineering, 2009.

[6] . L. C. H. Yeh, I. C., "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," tech. rep., Expert Systems with Applications, 36(2), 2473-2480, 2009.

[7] I.-C. Yeh, "Uci machine learning repository, default of credit cards data set," 2009.

[8] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," tech. rep., Technical report, SRI International, 1998.

[9] Y. W. Guang Li, "A privacy-preserving classification method based on singular value decomposition," tech. rep., The International Arab Journal of Information Technology, 2011.

Appendix