PROJECT DESCRIPTION:

I have proposed an algorithm to preserve the privacy of the database during data mining. The proposed algorithm has less data distortion. The algorithm, thus ensures that both mining can be done effectively and at the same time preserves the privacy of the database. Implemented the proposed algorithm as a java code.

PROPOSED ALGORITHM:

1. Select the quasi identifier with the highest number of unique values say Am such that Am⊆ Ai,….,Aj.

2. Perform selective generalization on Am as described in points 2.1 to 2.2.

2.1. Let G1,…,Gn be groups such that tuples in each group have same value of Am. The tuples not in any group of G1,….,Gn are generalized.

2.2. For the tuples in G1,….,Gn we consider the remaining quasi identifiers of Ai,….,Aj.

For each group in G1 to Gn repeat step 2.2.1. For c in 1 to n in 2.2.1:

2.2.1. For each tuple in Gc repeat steps 2.3.1.1 to 2.3.1.2.

2.2.1.1. For a tuple ensure that it has at least one more tuple in the same group which should have all the quasi identifier values (Ai,….,Aj) same as it. If so go to step 2.2.1.Else go to step 2.2.1.2.

2.2.1.2. Generalize the tuple.

3. For each generalized tuple in PT repeat step 3.1 .

3.1. Select tuples which have unique quasi identifier set Ai,…., Aj.

4. Slice PT such that each sliced table contains highly correlated values. Let the sliced tables of PT be B1,….,Bk, such that k is the total number of sliced tables.

5. In the sliced tables select a table Bh in B1,….,Bk such that it has at least one quasi identifier.

6. Perform selective shuffling on the selected table Bh. This is done by shuffling the tuples selected in step 3.

CODE DESCRIPTION:

I have just written the java code for the above algorithm with very few tuples. One can even extend this to more number of tuples to see how the algorithm works, in a detailed manner.