Data mining

Data Mining (DM) is the process of extracting useful information (e.g., rules) from large amount of data.

Data anonymity

Data are considered anonymous if you cannot link them to people. ID and quasi-ID (subset of public available attributes that can be used as ID) need to be masked or removed.

Can data mining results violate anonymity of individuals?

Surprisingly yes, they can!

How to find all anonymity breaches in DM results?

We developed naïve and optimized inference channel detectors that exploit theoretical results on closed sets.

Enforcing Pattern Anonymity

ADD and SUP algorithms can be used to block anonymity threats, by merging inference channels and then modifying the original support of patterns.

ADD increments the support of infrequent patterns, while SUP suppresses the information about infrequent data.

Both are shown to preserve information while removing anonymity threats.

Mining results from anonymized database can be useless...

...while pattern anonymization preserves information!