Instructions:

This is an instruction for AutoML users. To re-open this instruction, click “Help – AutoML User Instructions”.

[Step 1] (Optional) Change Attribute Types: Attributes with hierarchy files are set as “Quasi-identifying” as default; attributes without hierarchy files are set as “Insensitive” as default.

Steps to change attribute types:

* Select the “Configure transformation” tab
* Select a column under “Input data” tab on the left
* Find the “Data transformation” tab on the right, and change the field “Type”.

[Step 2] Add Privacy Models:

* Select the “Cofigure transformation” tab
* Find the “Privacy models” tab on the right bottom corner
* Click on the “+” to add a privacy model
* Select a privacy model and double click the knob to change parameter values
* Click “OK” to add the model
* Use “+” to add more privacy models, or use “-” to delete existing privacy models

[Step 3] Start Anonymization:

* Click on the green check at the top to start anonymization
* Wait for the anonymization progress bar to finish

[Step 4] Export anonymized data to AutoML:

* Select the “Analyze/enhance utility” tab, the output data is shown in the window on the right side
* Select “File – Export to AutoML” in the menu on the top, the output data will be stored in the path provided by AutoML

[Step 5] (Optional) Test on different privacy models:

* To try a different privacy model separately, repeat Step 2 to Step 4

[Step 6] Exit and go back to AutoML:

* Select “File – Exit” in the menu or click the cross sign on the corner of the window

Privacy Models:

[(e,d)-Differential privacy] Protects information about the individuals, but preserves information about the population.

Recommended Setting: e <= 1, d <= 1/size of dataset. Smaller e and d gives stronger privacy protection.

[k-Anonymity] Protects against identity disclosure. That is, it prevents an attacker from re-identify the records in the dataset.

Recommended Setting: k > 1 / maximum acceptable re-identification risk. Larger k gives stronger privacy protection.

[k-Map] A variant of k-anonymity, which considers explicit information about the underlying population.

Recommended Setting: k > 1 / maximum acceptable re-identification risk. Larger k gives stronger privacy protection.

[Average-reidentification-risk] Protects against marketer attacks (A re-identification attack that targets a large portion of the records.)

Recommended Setting: Acceptable re-identification risk.

[Population-uniqueness] Protects against identity disclosure in marketer attacks.

Recommended Setting: < 10% Low uniqueness (strong privacy protection), 10%~50% Medium uniqueness (medium privacy protection), >50% High uniqueness (low privacy protection)

[d-Presence] Protects against membership disclosure. That is, it prevents an attacker from inferring whether an individual’s record is included in the dataset.

Recommended Setting: d\_min >= 0, d\_max <= 0.5. Smaller range gives stronger privacy protection.

[l-Diversity] Protects against attribute disclosure. That is, it prevents an attacker from inferring the sensitive attributes in the dataset.

\*User needs to define the sensitive attributes under “Data Transformation” first.

Recommended Setting: l > 1 / acceptable probability of attribute disclosure. Larger l gives stronger privacy protection.

[t-Closeness] Protects against attribute disclosure. An alternative for l-Diversity.

\*User needs to define the sensitive attributes under “Data Transformation” first.

Recommended Setting: t <= 0.2. Smaller t gives stronger privacy protection.

[d-Disclosure privacy] A very strict measure for mitigating attribute disclosure.

\*User needs to define the sensitive attributes under “Data Transformation” first.

Recommended Setting: d <= 1.2. Smaller d gives stronger privacy protection.

[Sample-uniqueness] Records that are unique within the sample.

Recommended Setting: Smaller value gives stronger privacy protection.