Do-it-yourself: Data Generation with KNIME
Iris Adä

Modular Data Generation:
- Split data generation process into small and understandable modules.

Key Benefits:
- Visualization of generation process
- Easy modification
- Expandability
- Reusability
- Implemented as extension of KNIME

Overview:
- Modular Data Generation
  - Categorical
    - Conditional Label Assigner
    - Random Item Inserter
    - Random Label Assigner
    - One Rule Inserter
  - Misc
    - Empty Table Creator
    - One Row To Many
    - Random Boolean Assigner
    - Random Matcher
    - Stresser
    - TimeDelay Loop End
    - TimeDelay Loop Start
  - Numerical
    - Beta Distributed Assigner
    - Counter Generation
    - Gamma Distributed Assigner
    - Gaussian Distributed Assigner
    - Random Number Assigner
    - Time Series Generator

Association Rule Generation
- Generate baskets
- Increase support
- Apple, chips => gummy bears
- Find rules / frequent items

Generation of Complex Market Basket Data
- Customer Generation
- Name Distributed Assigner
- Double To Int
- One Row To Many
- One Row To Many
- Double To Int
- One Row To Many
- Get baskets
- Get baskets names
- Get baskets rows
- Get baskets
- Get basket item

Recursive Data Generation
- Generate a random walk
- Time Series Generation (New node)

Insert Simple Rules
- Simple rules based on multiple columns.

Missing Value Injection
- Inject missing values to random cells.

Split and Rejoin
- Modify only parts and rejoin results.

Apply Mathematical Functions
- e.g. create circular clusters

Cluster Generation
- Empty Table Creator
- Random Number Assigner
- Random Number Assigner
- Math Formula
- Create dimension
- Create variance
- Create y