Prof. Heitor Silvério Lopes
Prof. Thiago H. Silva

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CPGEI

Data Mining & Knowledge Discovery

Class 1d – Introduction to the Orange software

2025

What is the Orange datamining ?

- It is an opens-source data mining tool
- Developed in Python3 by the University of Ljubliana, Slovenia
  - https://orangedatamining.com/
- Orange uses the concept of "**visual programming**"
  - https://www.youtube.com/watch?v=H1ibqB_cvlE
- Current version: 3.38.1 (Windows, Linux)
- Created as a tool for Bioinformatics (currently, there are several add-ons)
- Detailed documentation:
  - https://orange3.readthedocs.io/projects/orange-data-mining-library/en/latest/
- It can work within the Anaconda Navigator enviroment or standalone
  - https://docs.anaconda.com/anaconda/navigator/install

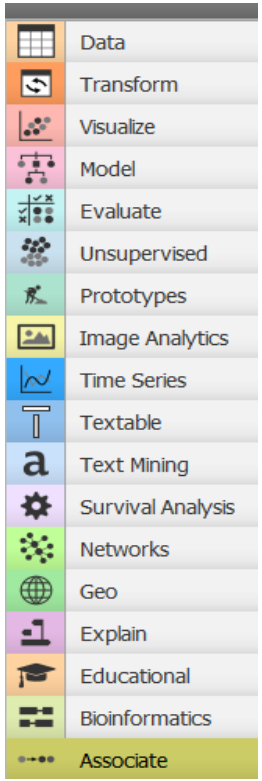
Orange structure

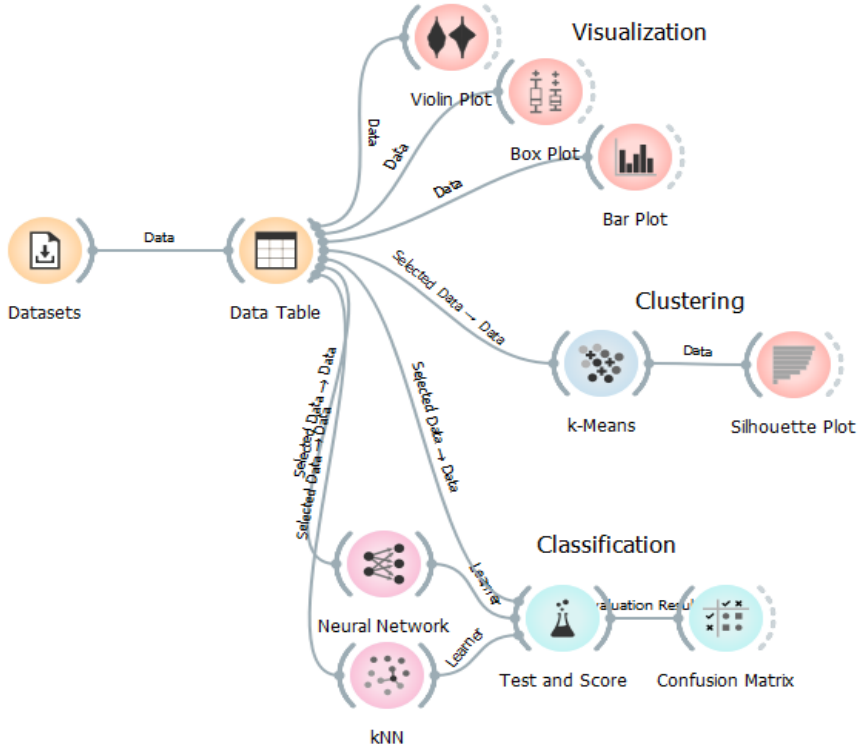
- All data manipulation and processing functions are accomplished by accessing functions represented by icons grouped by categories:
  - Basic:
    - Data
    - Transform
    - Visualize
    - Model
    - Evaluate
    - Unsupervised
    - Text Mining
    - Associate
  - Complimentary:
    - Bioinformatics
    - Single Cell
    - Educational
    - Geo
    - Image Analytics
    - Network
    - Time Series
    - Prototypes
    - Spectroscopy
    - Textable
    - Survival Analysis
    - World Happiness
    - Explain
    - Fairness

Orange visual programming

- Orange uses the concept of “visual programming” or “codeless machine learning”, that is, no need for coding

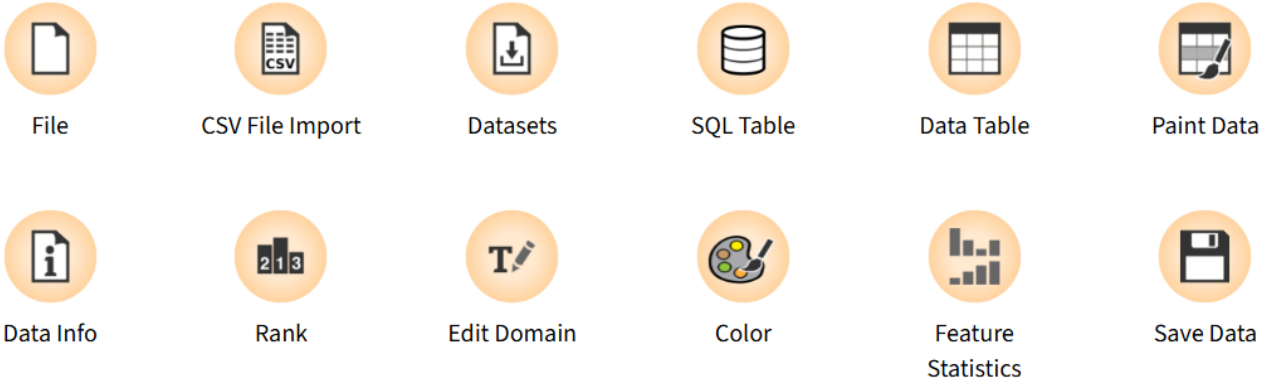
- A programming workflow is constructed by simply dragging icons into the canvas working area, and interconnecting them





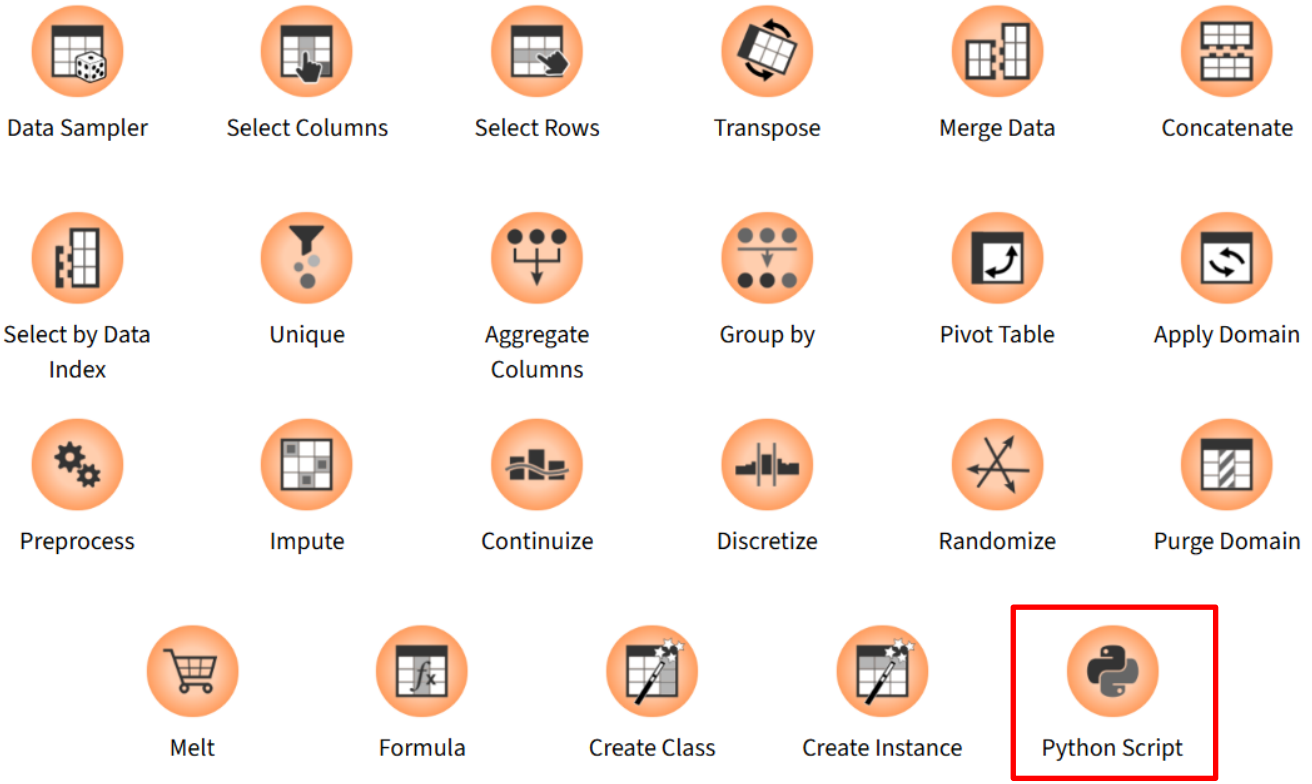
Orange functionalities

- Data
  - Reads data from the internet or files (tab, CSV, SQL)
  - Performs feature selection methods
  - Presents basic statistics of data
  - Constructs tables, saves selected data



Orange functionalities

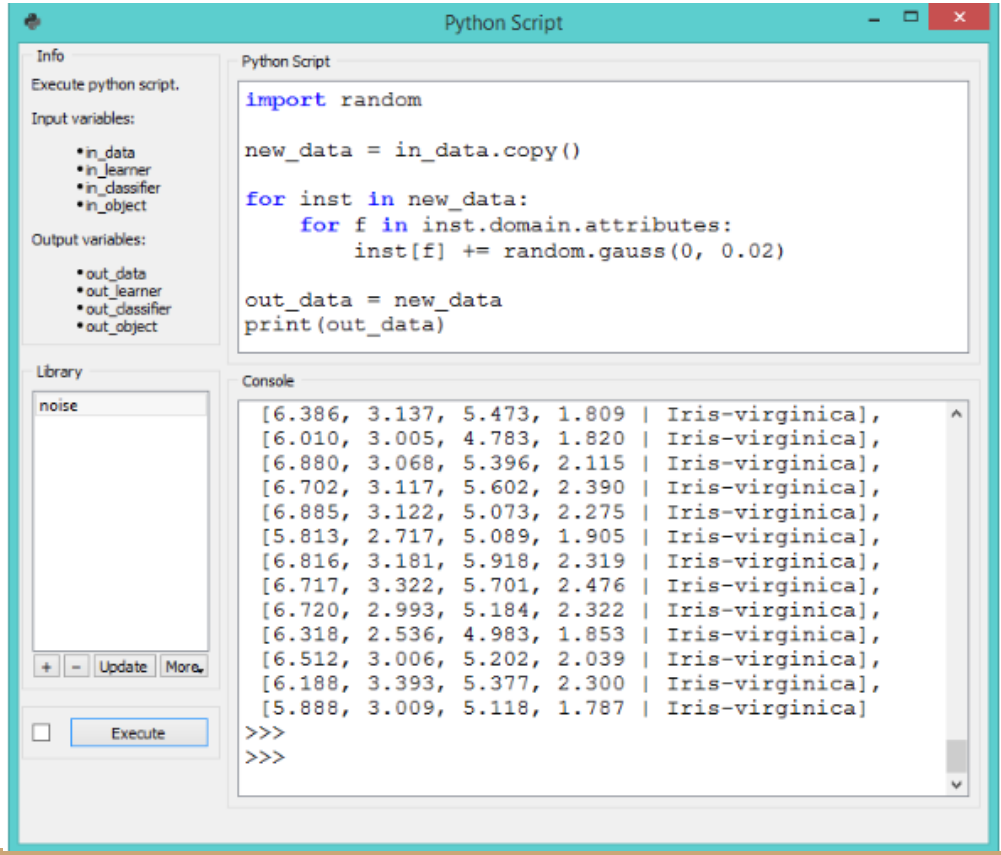
- Transform
  - Perform deveral data manipulations
  - Integration with user's Python scripts



Orange functionalities

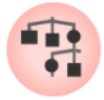
- This widget requires the use of standardized input/output classes

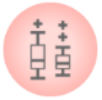
- Python scritpts example:
  - Adding Gaussian noite to input variables of a given dataset



Orange functionalities

- Visualize
  - Create several graphical representations of data
  - Rule viewer


Tree Viewer


Box Plot

Violin Plot

Distributions

Scatter Plot

Line Plot

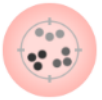
Bar Plot

Sieve Diagram

Mosaic Display

FreeViz

Linear Projection


Radviz

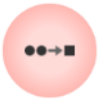
Heat Map

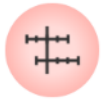
Venn Diagram

Silhouette Plot

Pythagorean Tree

Pythagorean Forest

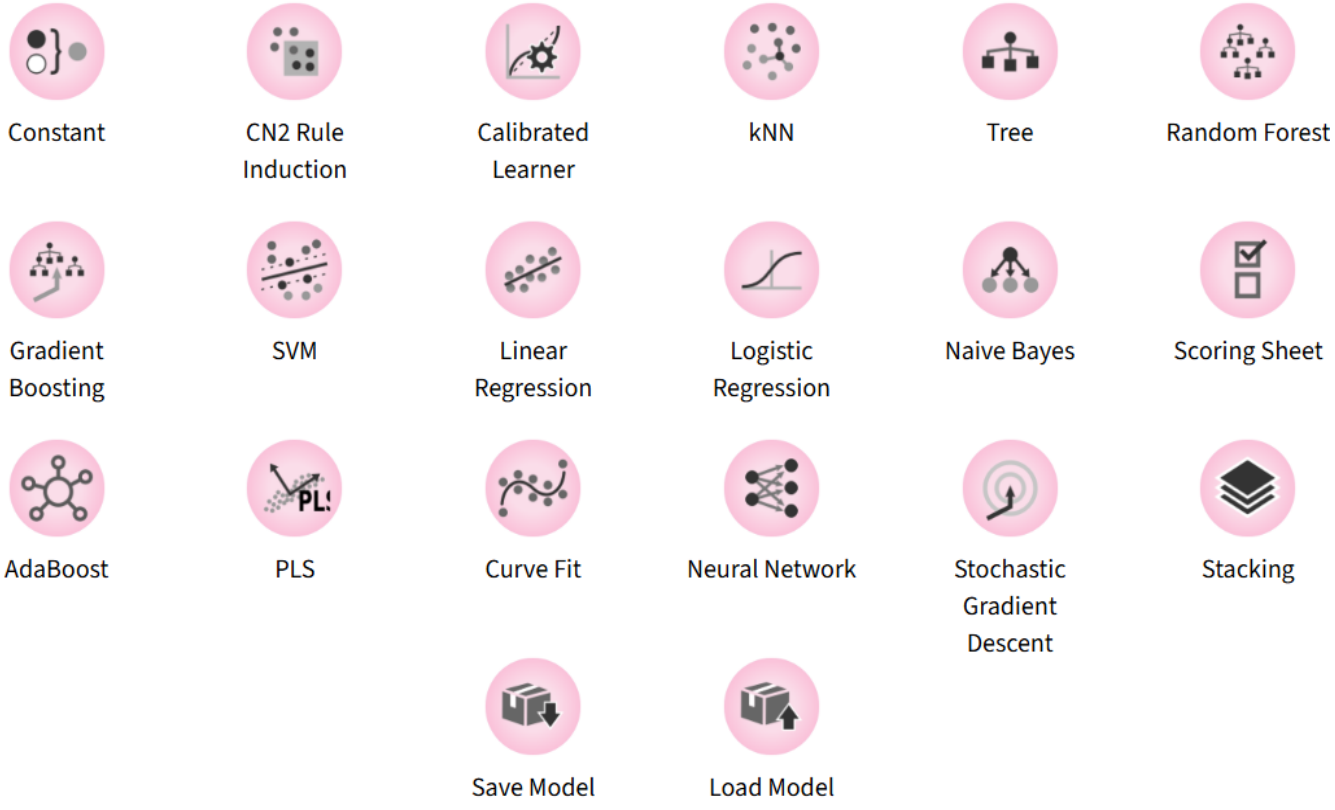

CN2 Rule Viewer


Nomogram

Scoring Sheet Viewer

Orange functionalities

- Model
  - Train/test machine learning models for classification, regression, curve fitting



Orange functionalities

- Evaluate
  - Evaluate the quality of the trained models

Test and Score

Predictions

Confusion Matrix

ROC Analysis

Performance Curve

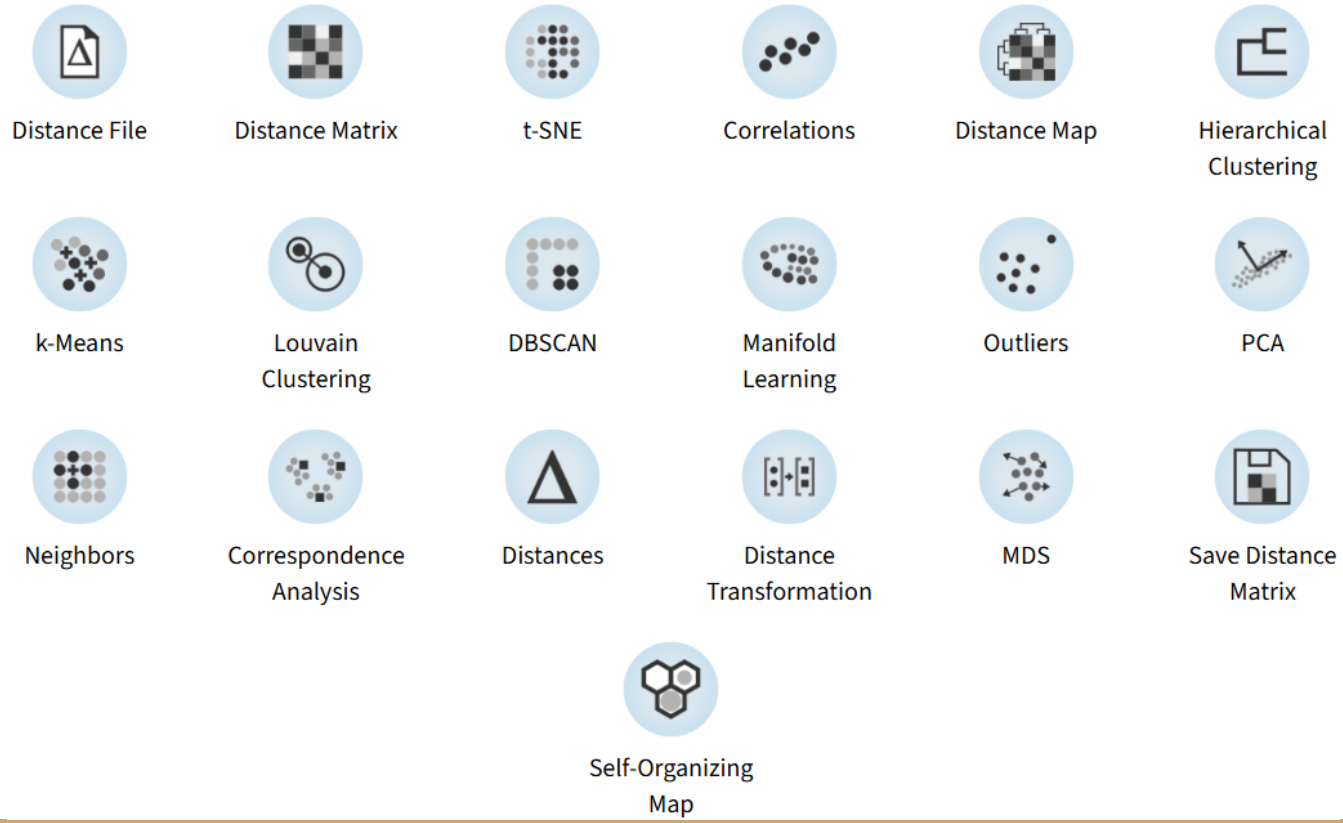
Calibration Plot

Permutation Plot

Parameter Fitter

Orange functionalities

- Unsupervised
  - Provide unsupervised methods for analyzing data
  - Methods for data clustering
  - Methods for dimensionality reduction

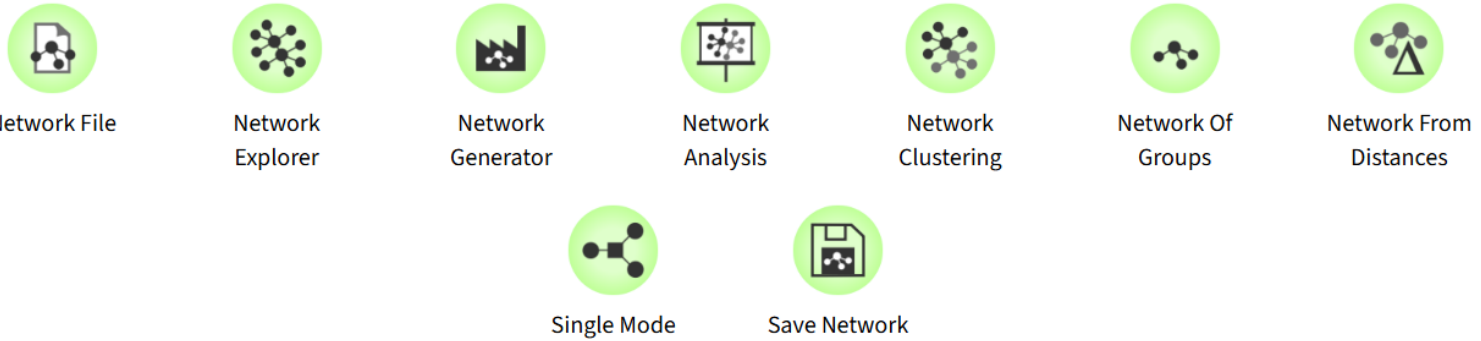


Orange functionalities

- Text mining
  - Collects data from several sources
  - Many tools for preprocessing texts
  - Many tools for text mining

Orange functionalities

- Network
  - Provide tools for dealing and processing graphs



Orange functionalities

- Image Analytics
  - Provide tools for dealing with images
  - Several pre-trained DNN for extracting features from images

Import Images

Image Viewer

Image Embedding

Image Grid

Save Images

Orange functionalities

- Associate
  - Tools for associative analysis (frequent patterns discovey)

Frequent Itemsets

Association Rules

Orange functionalities

- Application-specific groups of tools:
  - Spectroscopy
  - Survival Analysis
  - Bioinformatics
  - Single Cell
  - Networks
  - Geo
  - Educational
  - Time Series
  - Explain
  - Fairness