



Data Mining & Knowledge Discovery

Class 1d - Introduction to the
Orange software

2025

What is the Orange datamining ?



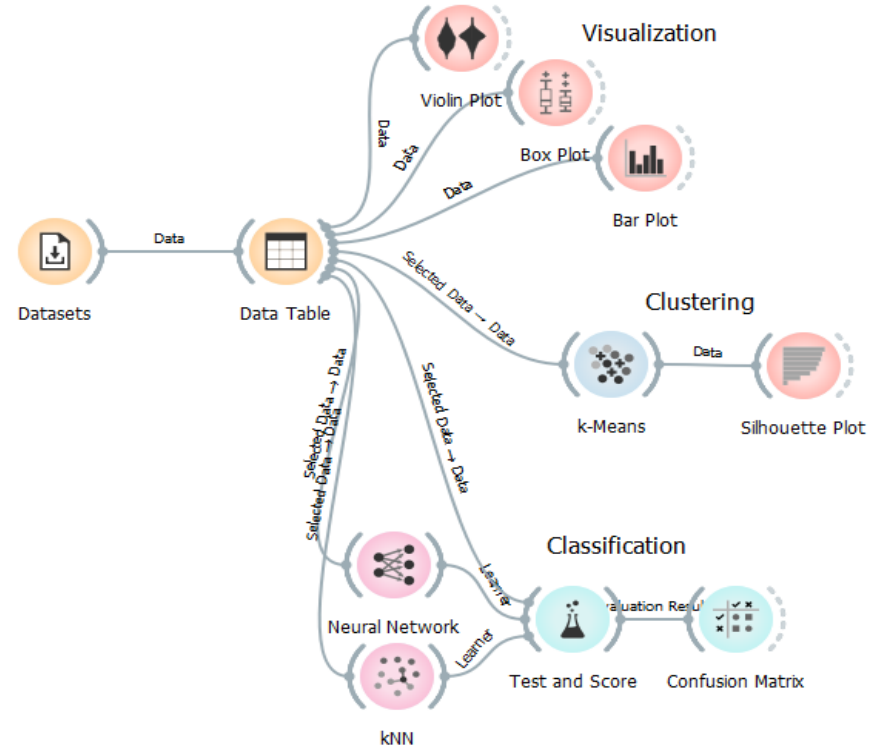
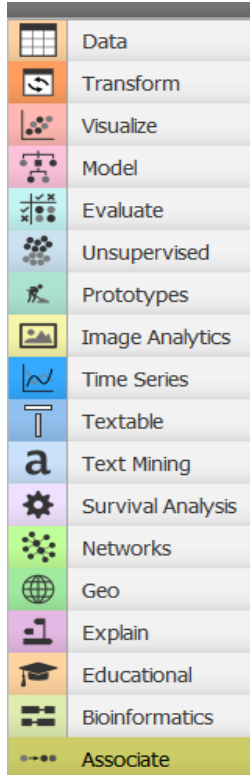
- It is an opens-source data mining tool
- Developed in Python3 by the University of Ljubliana, Slovenia
 - <https://orangedatamining.com/>
- Orange uses the concept of “**visual programming**”
 - https://www.youtube.com/watch?v=H1ibqB_cvIE
- Current version: 3.38.1 (Windows, Linux)
- Created as a tool 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



File



CSV File Import



Datasets



SQL Table



Data Table



Paint Data



Data Info



Rank



Edit Domain



Color



Feature Statistics



Save Data

Orange functionalities

- Transform

- Perform several data manipulations
- Integration with user's Python scripts



Data Sampler



Select Columns



Select Rows



Transpose



Merge Data



Concatenate



Select by Data Index



Unique



Aggregate Columns



Group by



Pivot Table



Apply Domain



Preprocess



Impute



Continue



Discretize



Randomize



Purge Domain



Melt



Formula



Create Class



Create Instance



Python Script

Orange functionalities

- This widget requires the use of standardized input/output classes
- Python scripts example:
 - Adding Gaussian noise to input variables of a given dataset

```
Python Script

import random

new_data = in_data.copy()

for inst in new_data:
    for f in inst.domain.attributes:
        inst[f] += random.gauss(0, 0.02)

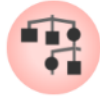
out_data = new_data
print(out_data)
```

```
>>>
[[6.386, 3.137, 5.473, 1.809 | Iris-virginica],
 [6.010, 3.005, 4.783, 1.820 | Iris-virginica],
 [6.880, 3.068, 5.396, 2.115 | Iris-virginica],
 [6.702, 3.117, 5.602, 2.390 | Iris-virginica],
 [6.885, 3.122, 5.073, 2.275 | Iris-virginica],
 [5.813, 2.717, 5.089, 1.905 | Iris-virginica],
 [6.816, 3.181, 5.918, 2.319 | Iris-virginica],
 [6.717, 3.322, 5.701, 2.476 | Iris-virginica],
 [6.720, 2.993, 5.184, 2.322 | Iris-virginica],
 [6.318, 2.536, 4.983, 1.853 | Iris-virginica],
 [6.512, 3.006, 5.202, 2.039 | Iris-virginica],
 [6.188, 3.393, 5.377, 2.300 | Iris-virginica],
 [5.888, 3.009, 5.118, 1.787 | Iris-virginica]]
>>>
```

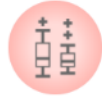
Orange functionalities

- Visualize

- Create several graphical representations of data
- Rule viewer



Tree Viewer



Box Plot



Violin Plot



Distributions



Scatter Plot



Line Plot



Bar Plot



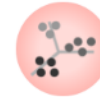
Sieve Diagram



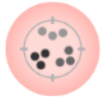
Mosaic Display



FreeViz



Linear
Projection



Radviz



Heat Map



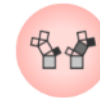
Venn Diagram



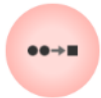
Silhouette Plot



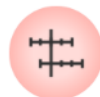
Pythagorean
Tree



Pythagorean
Forest



CN2 Rule Viewer



Nomogram



Scoring Sheet
Viewer

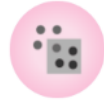
Orange functionalities

- Model

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



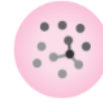
Constant



CN2 Rule Induction



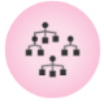
Calibrated Learner



kNN



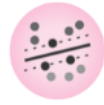
Tree



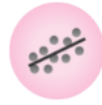
Random Forest



Gradient Boosting



SVM



Linear Regression



Logistic Regression



Naive Bayes



Scoring Sheet



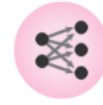
AdaBoost



PLS



Curve Fit



Neural Network



Stochastic Gradient Descent



Stacking



Save Model



Load Model

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



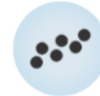
Distance File



Distance Matrix



t-SNE



Correlations



Distance Map



Hierarchical Clustering



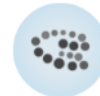
k-Means



Louvain Clustering



DBSCAN



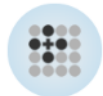
Manifold Learning



Outliers



PCA



Neighbors



Correspondence Analysis



Distances



Distance Transformation



MDS



Save Distance Matrix



Self-Organizing Map

Orange functionalities

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



Corpus



Import Documents



Create Corpus



The Guardian



NY Times



Pubmed



Twitter



Wikipedia



Preprocess Text



Corpus to Network



Bag of Words



Document Embedding



Similarity Hashing



Sentiment Analysis



Tweet Profiler



Topic Modelling



LDAvis



Corpus Viewer



Score Documents



Word Cloud



Concordance



Document Map



Word Enrichment



Duplicate Detection



Word List



Extract Keywords



Annotated Corpus Map



Ontology



Orange functionalities

- Network
 - Provide tools for dealing and processing graphs



Network File



Network Explorer



Network Generator



Network Analysis



Network Clustering



Network Of Groups



Network From Distances



Single Mode



Save Network

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 discovery)



Frequent
Itemsets



Association
Rules

Orange functionalities

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