MachineLearnAthon - Microlecture
No code

MachineLearnAthon
A project Co-funded by the Erasmus+ programme of the European Union

No-Code approach

- Minimal coding required, visual interface dominate .
- Designed for business users and non-technical staff.
- Focus on accessibility and ease of use.

Key features:

- Visual tools and drag-and-drop interfaces.
- Pre-built templates for rapid development.
- Minimal technical expertise required.

Use Cases:

- Ideal for small-scale projects and prototyping.
- Common in business intelligence and automation.
- Used to create dashboards, forms, and reports.

Orange Data Mining

- An open-source data mining and machine learning platform.
- Provides a visual programming interface for data analysis.
- Designed for both beginners and experts in data science.

Origins of Orange

- Developed by the University of Ljubljana:
- Created by the Laboratory of Artificial Intelligence at the University of Ljubljana, Slovenia.
- Early development began in the late 1990s as a tool for teaching and research.

Vision

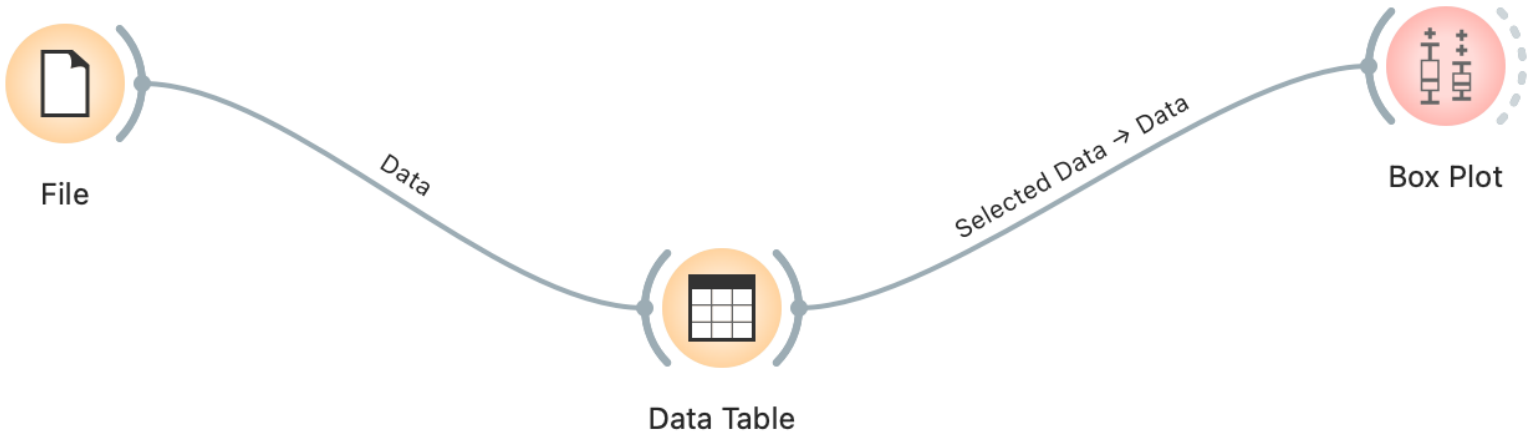
- To democratize access to advanced data mining tools.
- Combine ease of use with powerful analytical capabilities.

Download

- https://orangedatamining.com/

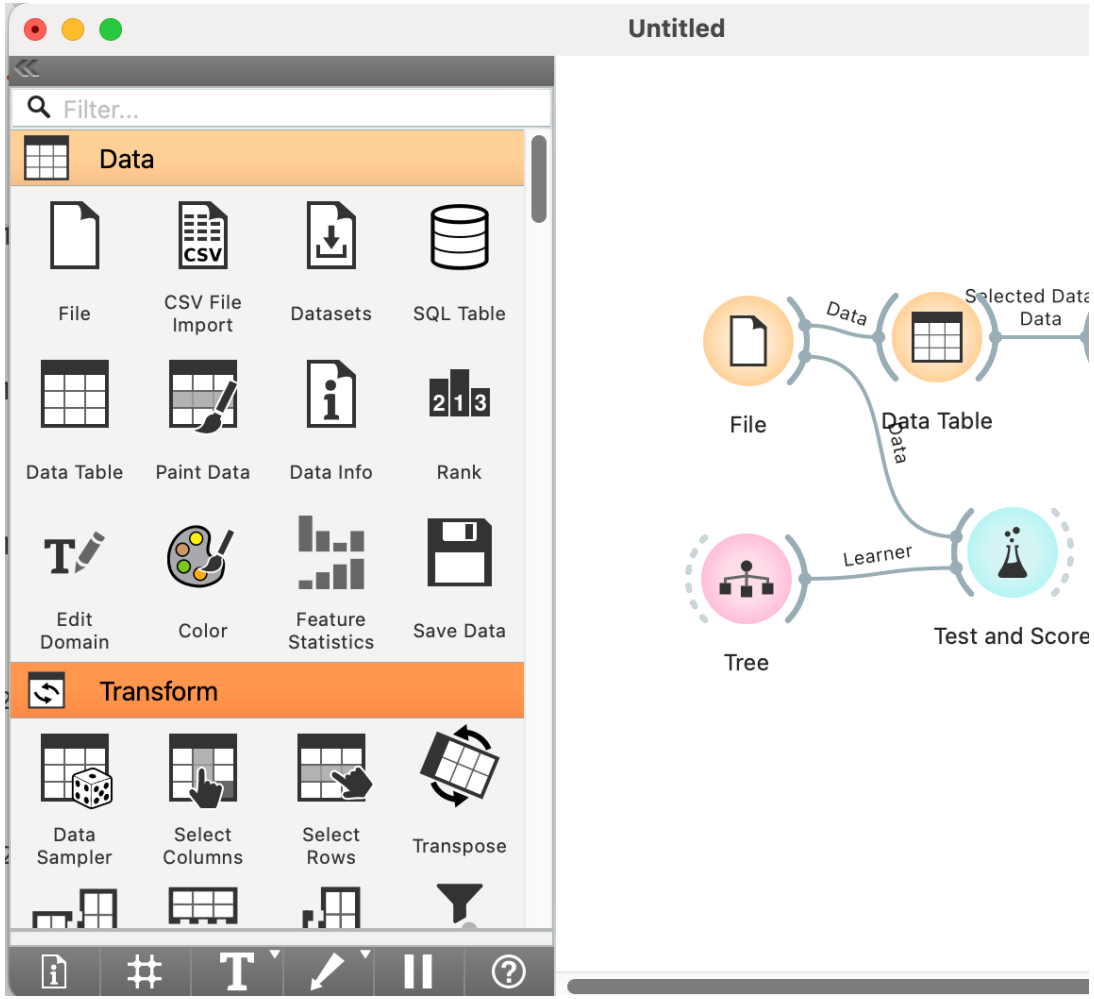
Orange ML: Overview and Features

- Visual programming for ML tasks.
- Interactive widgets for data analysis and modeling.
- Extensive libraries for preprocessing and visualization.
- Modular design: Add-ons for advanced tasks (e.g., text mining).
- Ease of use: Ideal for education, research, and prototyping.



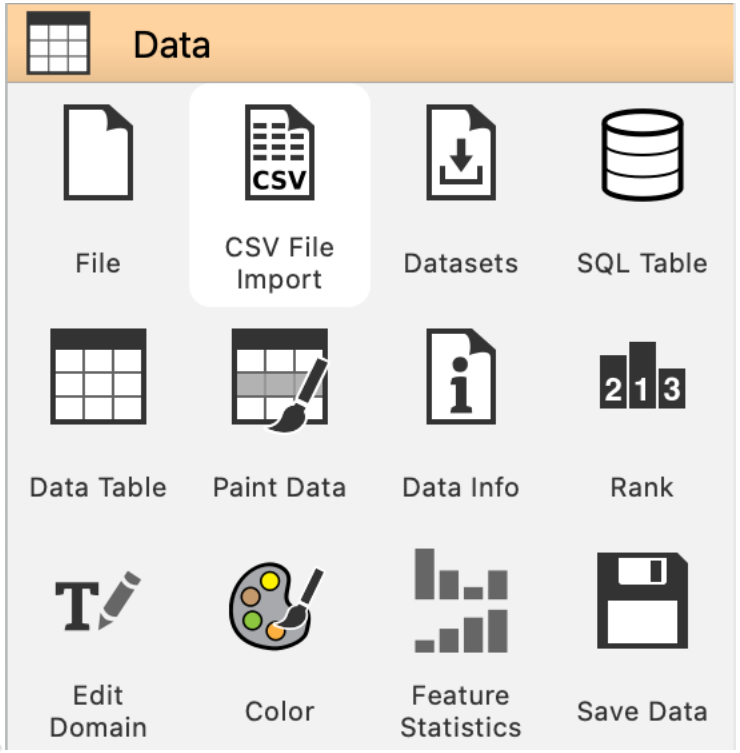
Orange Workflow

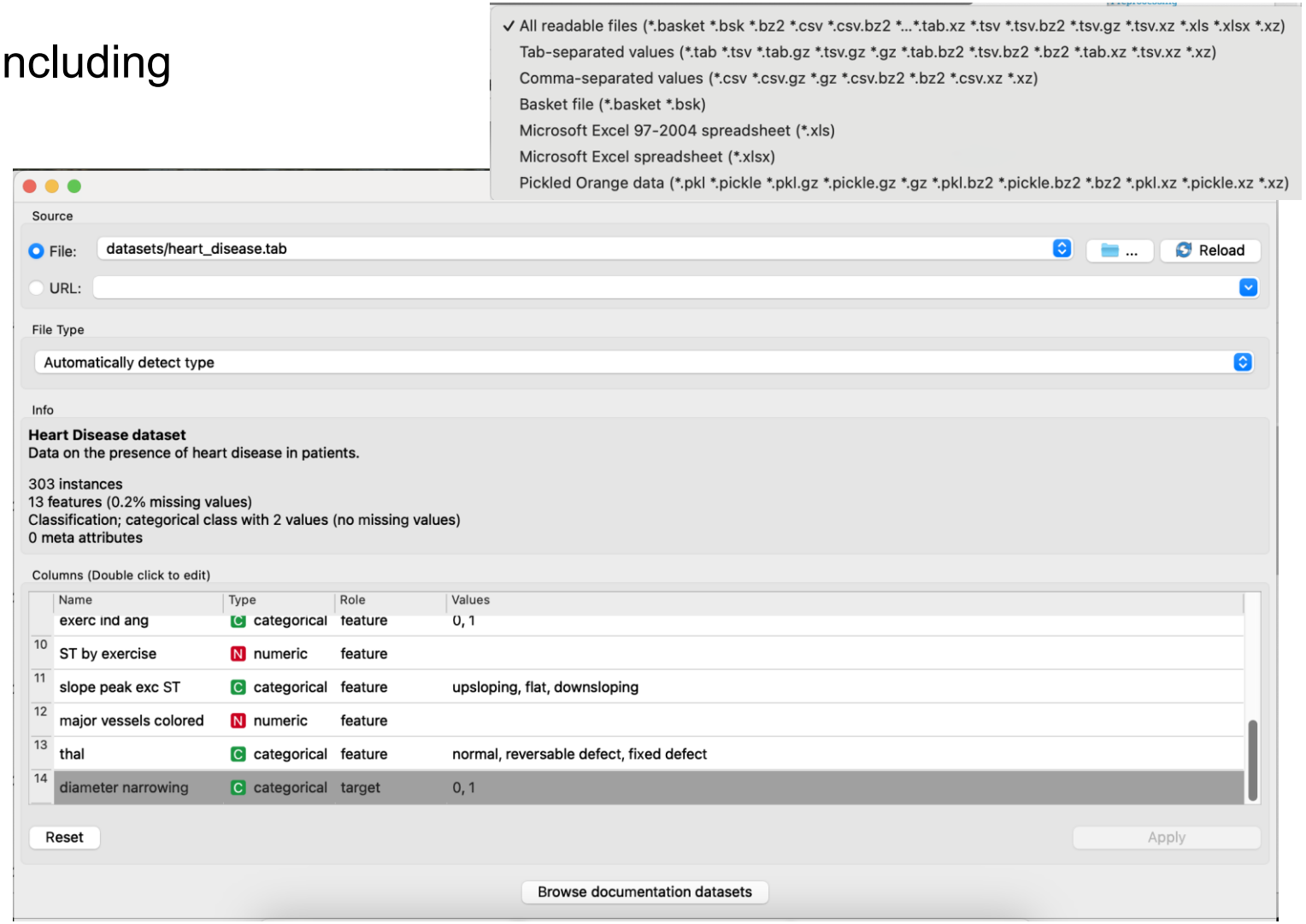
- Orange workflows consist of interconnected widgets.
- Widgets perform tasks such as data preprocessing, modeling, and visualization.
- Workflows are modular and highly customizable.



Step 1: Data Loading and Preprocessing

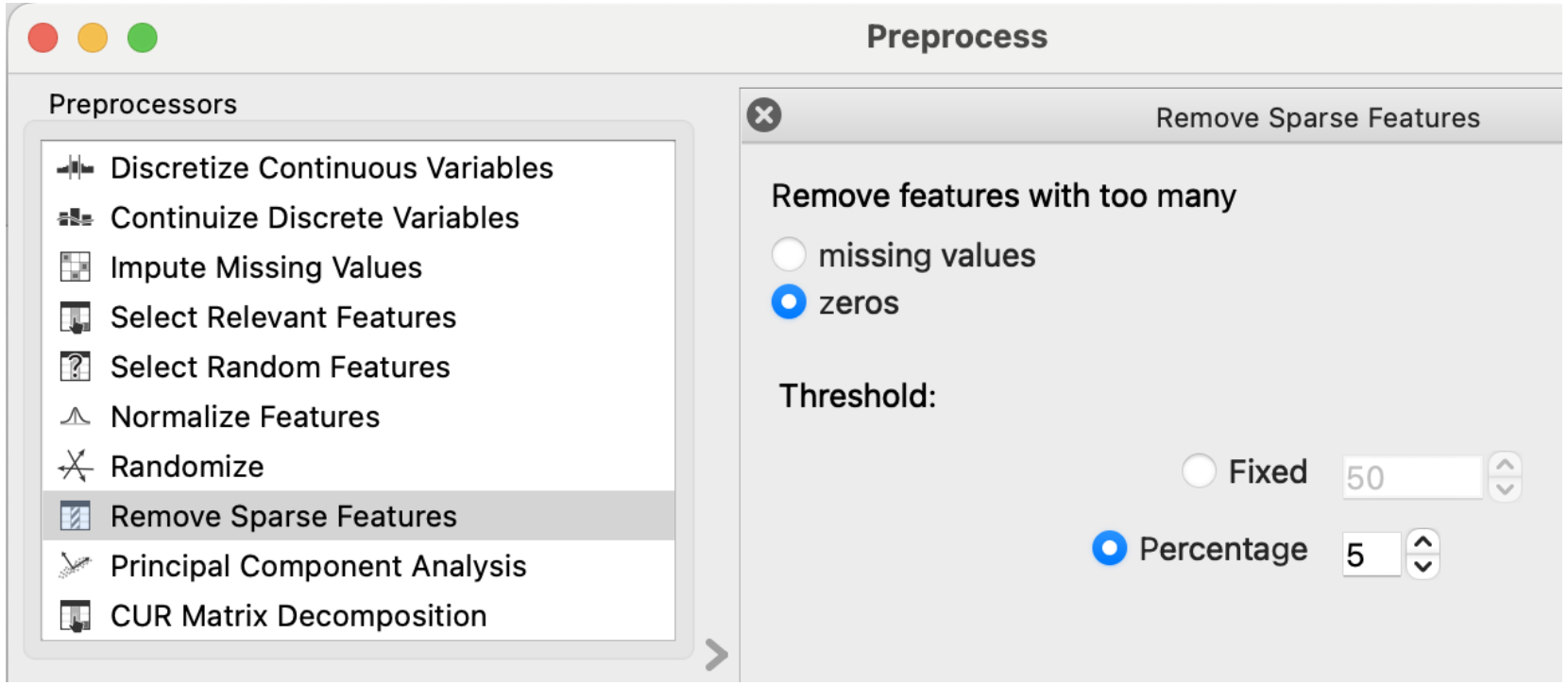
- Supports various data formats including CSV, Excel, and SQL

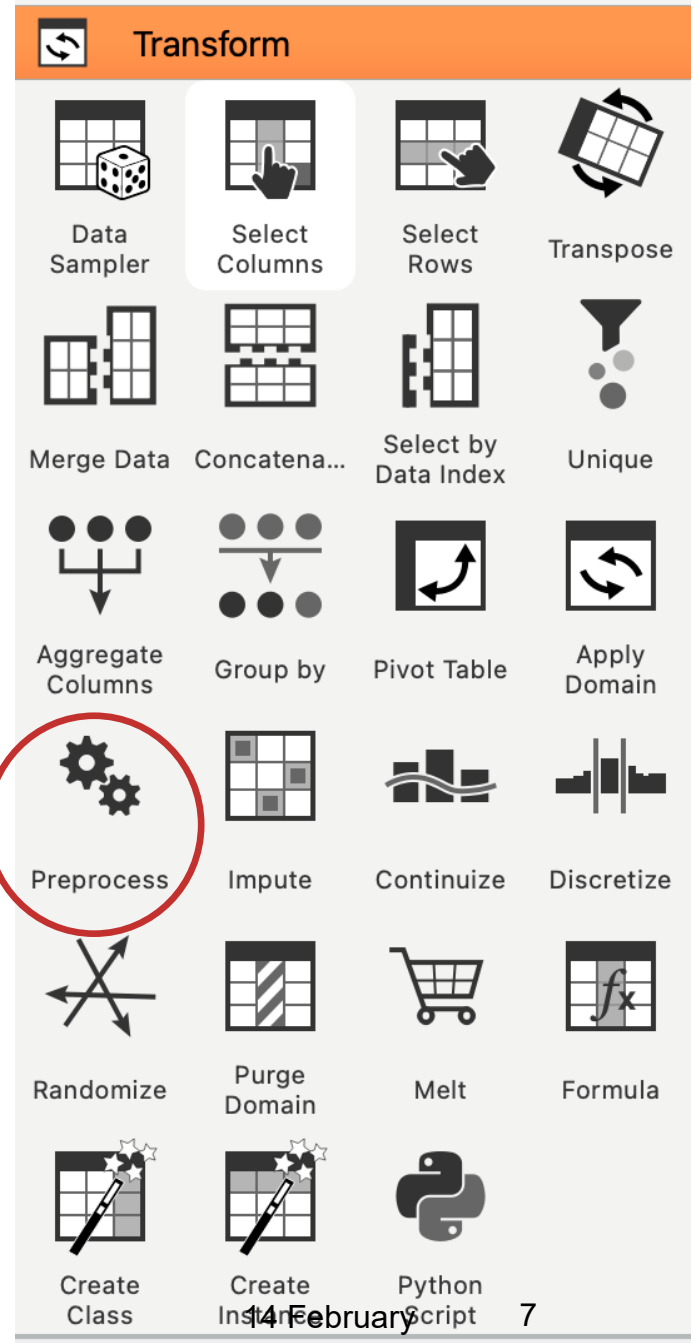




Step 2: Transformation

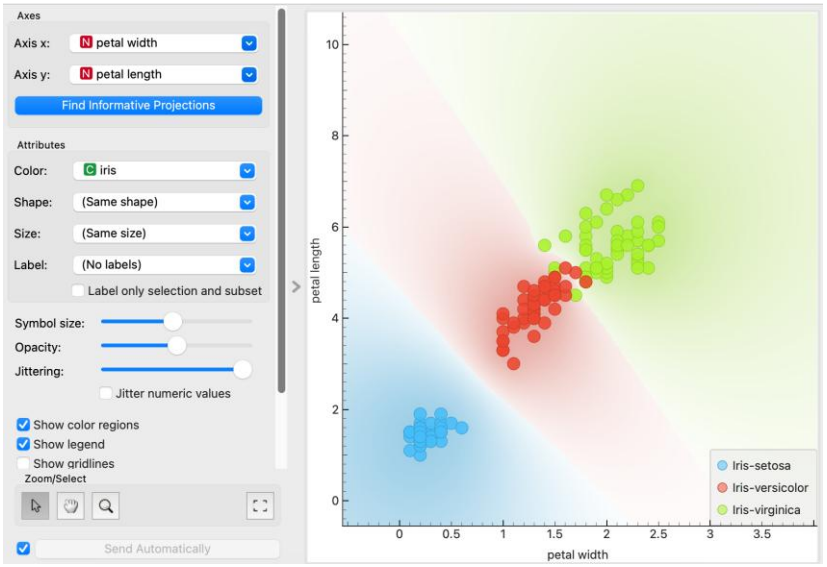
- Widgets for data cleaning, feature selection, and transformation.
- Handles missing values and data imputation.

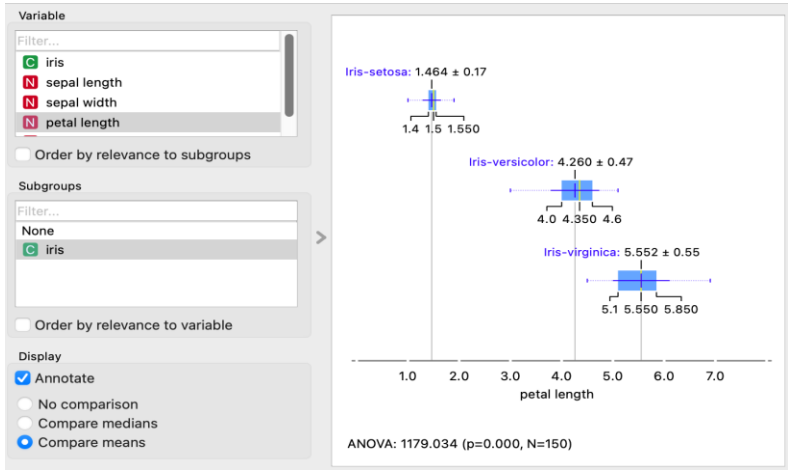


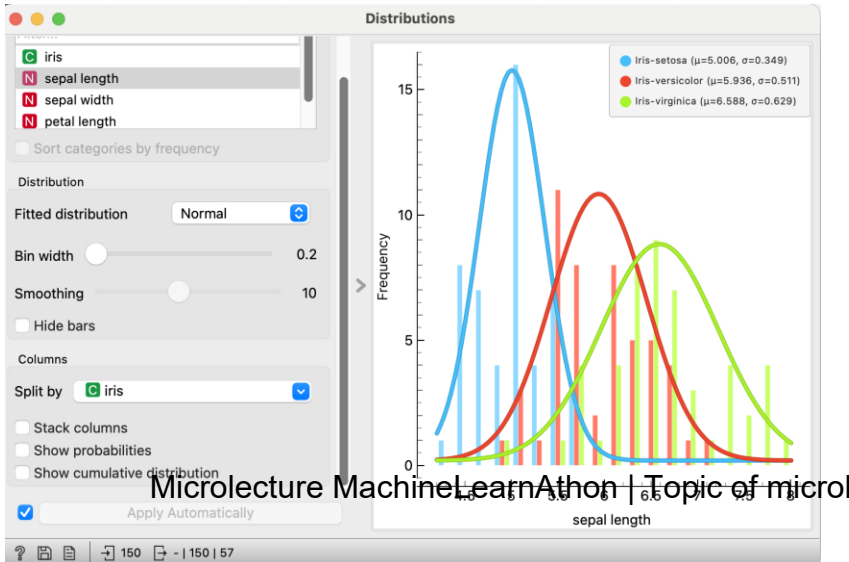


Step 3: Visualization

- Provides interactive plots for data exploration.
- Helps identify patterns and outliers in the data.

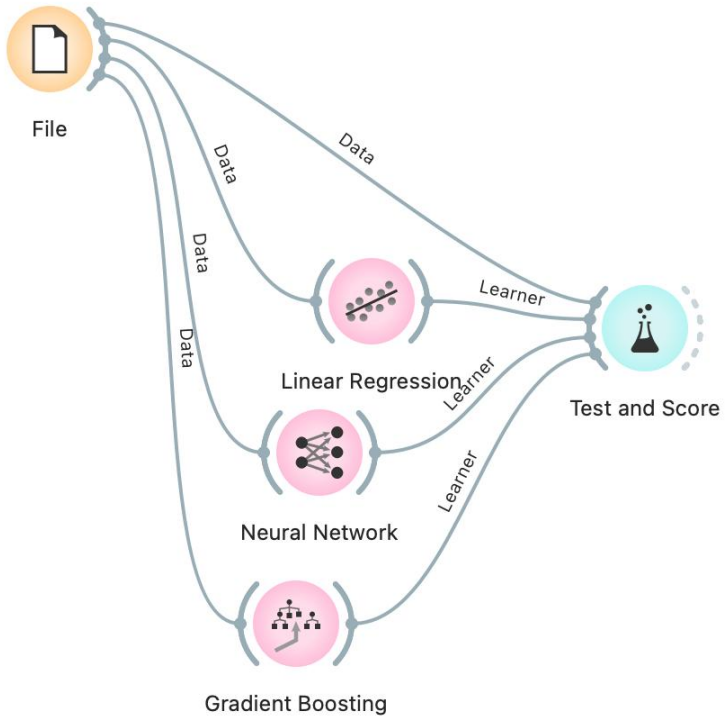


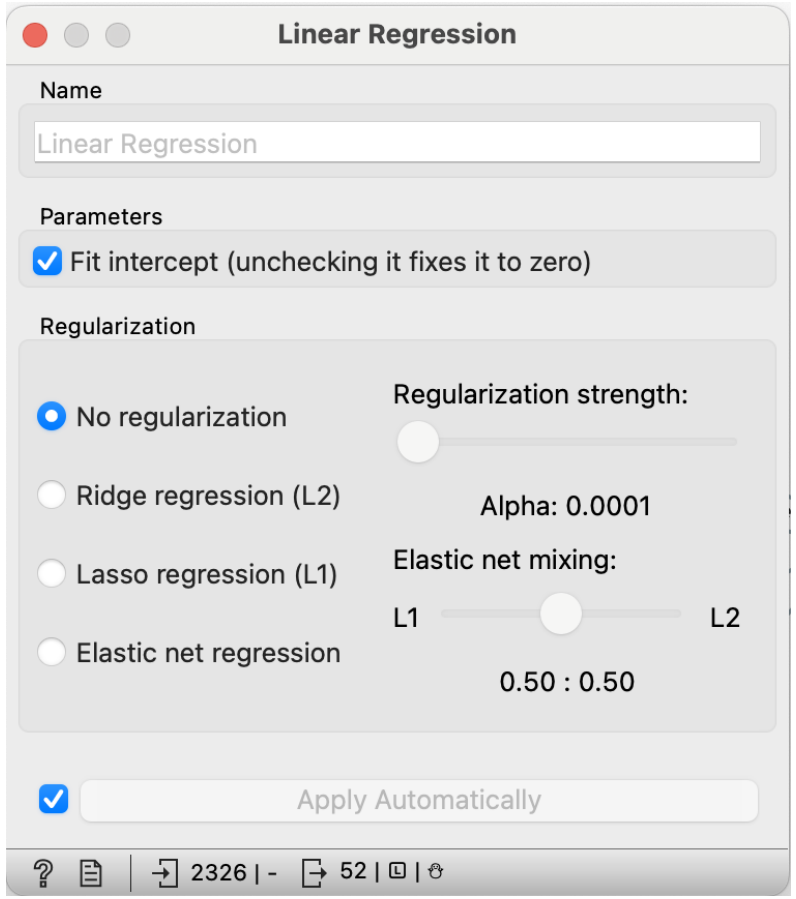


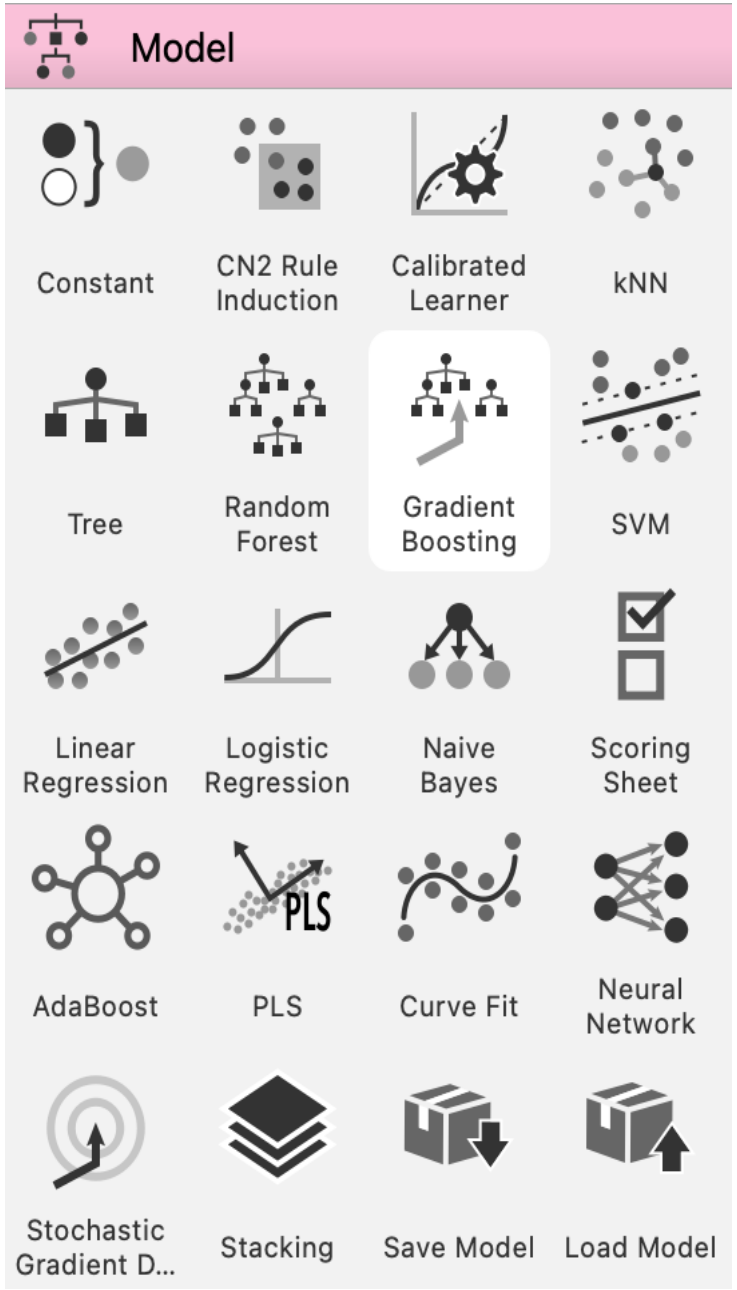


Step 4: Model Building

- Supports classification, regression, and clustering.



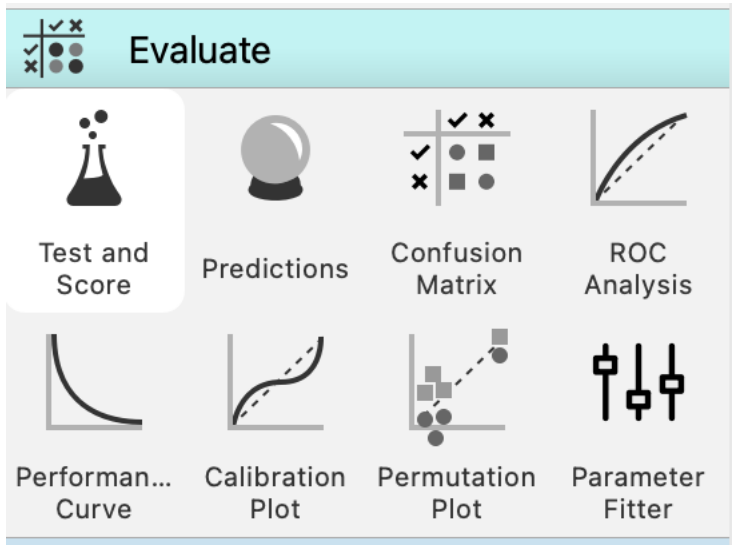


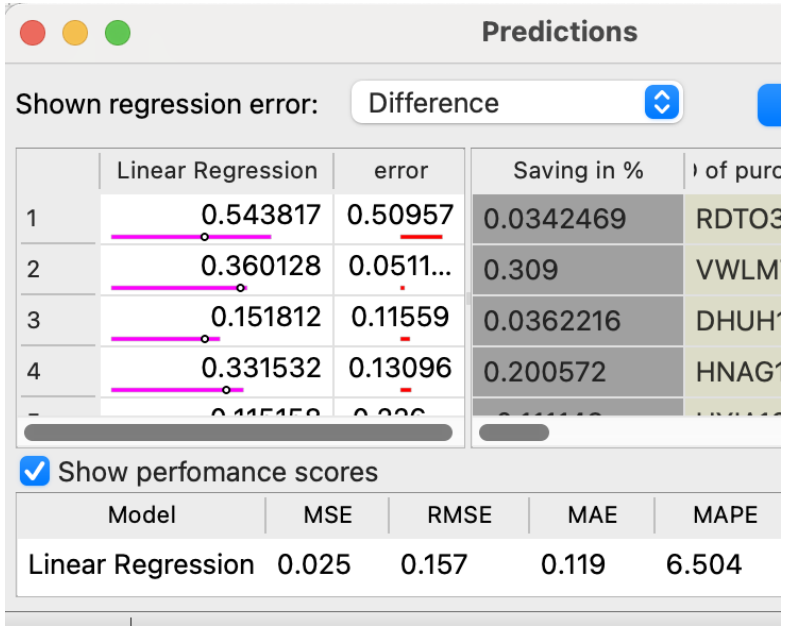


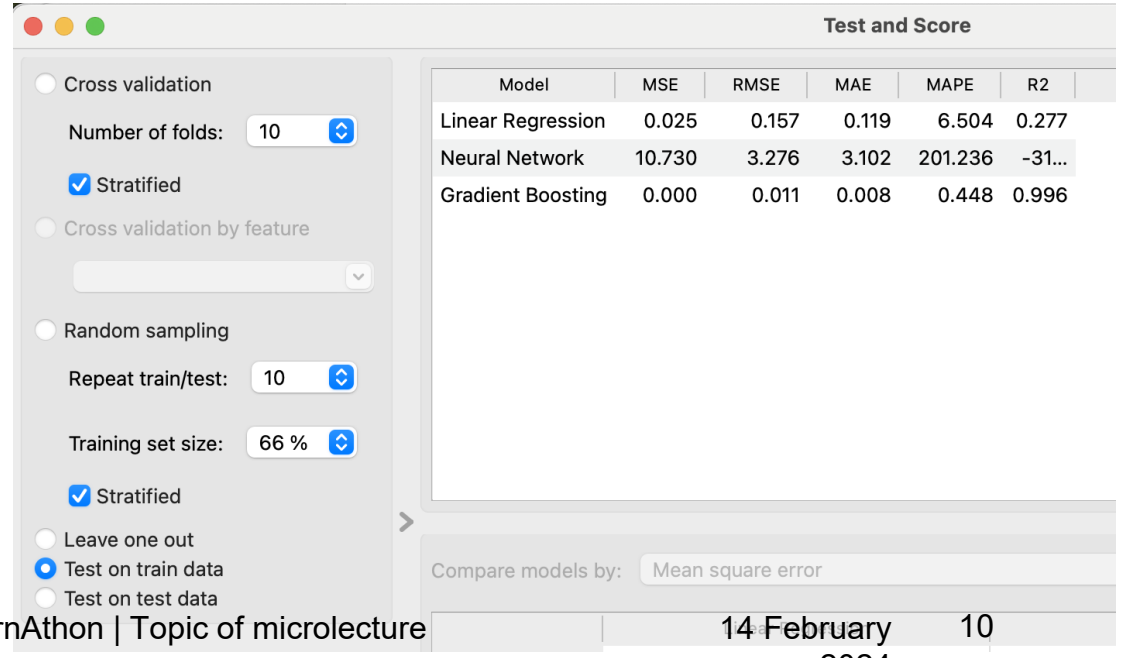
technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Step 5: Model Evaluation

- Evaluation widgets calculate metrics like accuracy, precision, and recall.
- Supports cross-validation and ROC curve analysis.
- Provides visual feedback on model performance.

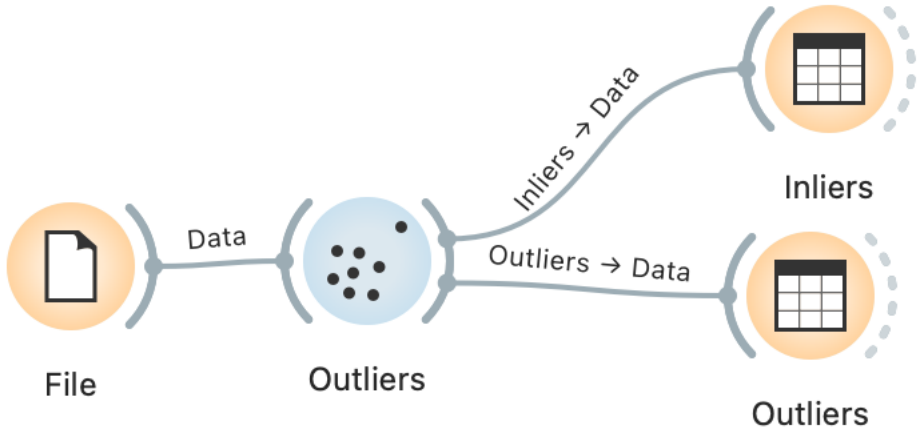


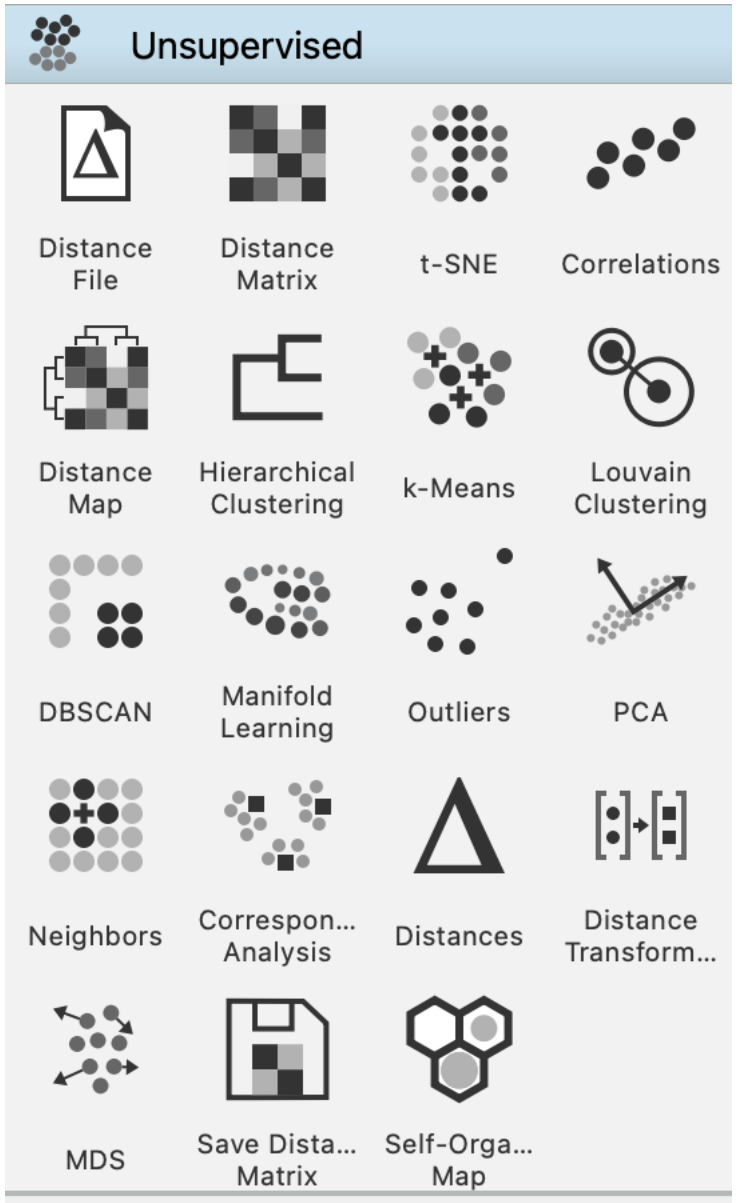




Unsupervised

- Unsupervised learning focuses on finding hidden patterns or structures in unlabeled data.
- Widget like clustering, dimensionality reduction, and data visualization.



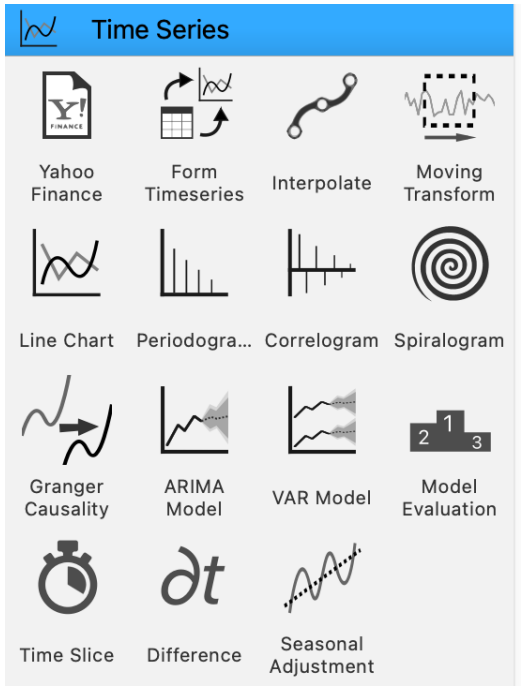


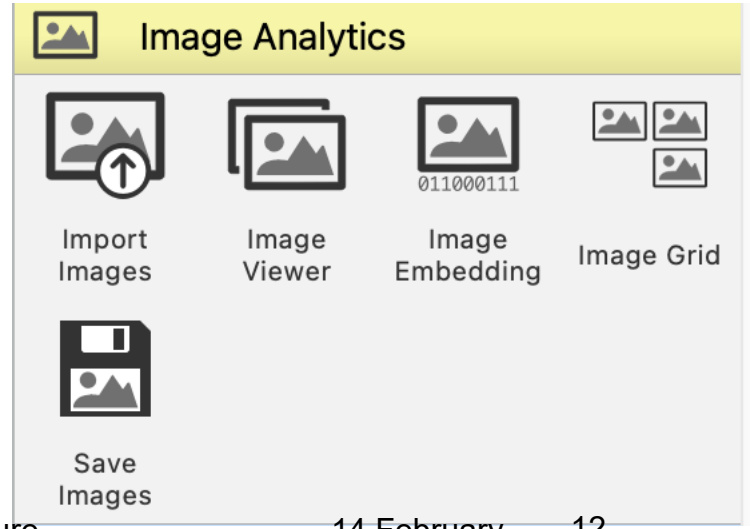
Advanced Widgets in Orange

- Text Mining: For natural language processing tasks.
- Image Analytics: For processing and analyzing images.
- Time Series: For analyzing sequential data.

	Name	Version	Action
☑	Orange3	3.38.0	
☐	Associate	1.3.0	
☐	Bioinformatics	4.8.4	
☐	Educational	0.8.0	
☐	Explain	0.6.10	
☐	Fairness	1.0.1	
☐	Geo	0.5	
☐	Image Analytics	0.13.0	
☐	Network	1.10	
☐	Prototypes	0.21.2	

	Name	Version	Action
☐	Prototypes	0.21.2	
☐	Pumice	1.0	
☐	Single Cell	1.7.2	
☐	Spectroscopy	0.7.2	
☐	Survival Analysis	0.6.1	
☐	Text	1.16.1	
☐	Textable	3.1.11	
☐	Timeseries	0.6.3	
☐	World Happiness	0.1.12	
☐	SNOM	0.0.4	





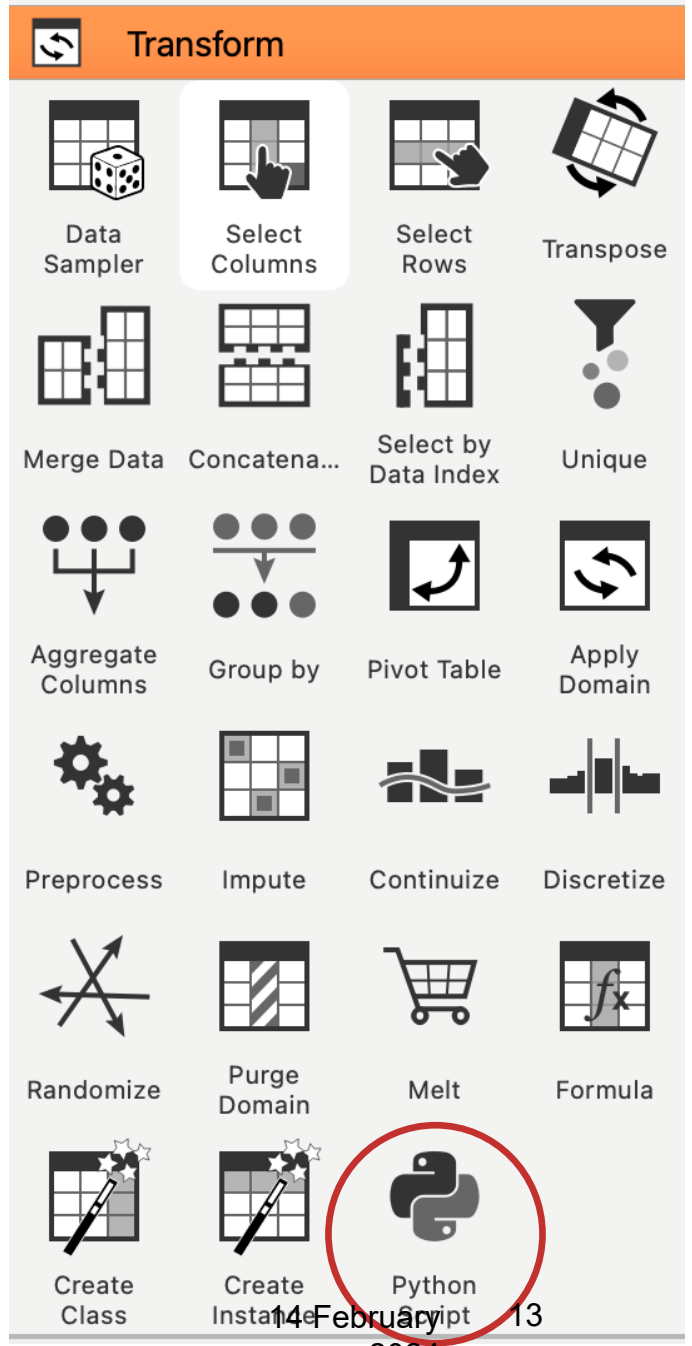
technische universität dortmund
TU Delft
Co-funded by the Erasmus+ programme of the European Union
Machine LearnAthon

Core Libraries in Orange

- **NumPy and SciPy:** For numerical and scientific computing.
- **Scikit-learn:** Backend for machine learning algorithms.
- **Matplotlib:** For visualization support.

Comparison with other Low code /No code tools

- Compared to **KNIME** and **RapidMiner,** Orange focuses on simplicity.
- More user-friendly than **Weka** for beginners.
- Less flexible than Python libraries like **Scikit-learn.**



Thank you very much for your attention!

Literature

- Alamin, M. A. A., & Uddin, G. (2018). Challenges and barriers of using low code software for machine learning. *ACM Transactions*.
- Data Valley. (n.d.). *Kaggle survey insights and analysis*. Retrieved December 1, 2024, from https://datavalley.technology/kaggle/
- Hasim, N., & Haris, N. A. (2015). A study of open-source data mining tools for forecasting. *Proceedings of the ACM International Conference*. https://doi.org/10.1145/2701126.2701152
- Kaggle. (2021). *Kaggle State of Machine Learning and Data Science Report 2021*. Retrieved from https://storage.googleapis.com/kaggle-media/surveys/Kaggle%20State%20of%20Machine%20Learning%20and%20Data%20Science%20Report%202021.pdf
- Kaggle. (2022). *Kaggle State of Machine Learning and Data Science Report 2022*. Retrieved from https://storage.googleapis.com/kaggle-media/surveys/Kaggle%20State%20of%20Machine%20Learning%20and%20Data%20Science%20Report%202022.pdf

Literature (cont.)

- Naik, A., & Samant, L. (2016). Correlation review of classification algorithm using data mining tools: WEKA, Rapidminer, Tanagra, Orange, and Knime. *Procedia Computer Science, 85,* 662-668. https://doi.org/10.1016/j.procs.2016.05.251
- Ng Mh. (2024). A comparison of the top 3 tools for machine learning beginners: Orange vs PyCaret vs Scikit. *Tinkercademy Build Log*. Retrieved from https://blog.tinkercademy.com/orange-v-s-pycaret-v-s-scikit-a-comparison-of-beginner-machine-learning-libraries-2d79c0e43e3f
- Rexer Analytics. (2023). *Data Science Survey 2023: Highlights from the TechCast (July 27, 2023)*. Retrieved December 1, 2024, from https://andouc.org/wp-content/uploads/2023/10/DSS-2023-Highlights-REXER-TechCast-7-27-23-CDChanges.pdf
- Wahbeh, A. H., Al-Radaideh, Q. A., Al-Kabi, M. N., & Al-Shawakfa, E. (2011). A comparison study between data mining tools over some classification methods. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/SpecialIssue.2011.010304
- Yan, Z. (2022). The impacts of low/no-code development on digital transformation and software development. *University of Toronto*. Retrieved from https://arxiv.org/pdf/2112.14073v1.pdf
- Yeh, I.-C. (2016). *Default of credit card clients dataset*. UCI Machine Learning Repository. University of California, Irvine. https://doi.org/10.24432/C55S3H

The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.