

## Mirco lecture Glossary

This document provides a glossary of topics covered in the MachineLearnAthon micro-lectures. For each topic, it specifies the corresponding micro-lecture and slide numbers where further information can be found. All entries are arranged alphabetically to ensure easy navigation and quick reference for learners and educators.

Topic	Micro Lecture	Slide
<b>Local Interpretable Model-Agnostic Explanation (LIME)</b>	Interpretability part 4	5
<b>"Best" straight line with given values</b>	Simple and multivariate linear regression	4
<b>(S)ARIMA</b>	Forecasting Methods	4
<b>Accumulated Local Effects (ALE) Plot</b>	Interpretability part 3	5-7
<b>Advanced Widgets in Orange</b>	No Code	12
<b>Advantages</b>	LightGBM	4
<b>Analysis of the impact of the size and number of rooms on the price of the apartment</b>	Simple and multivariate linear regression	12, 13
<b>Application example</b>	Random Forest	7
<b>Applications of neural networks for regression</b>	Neural Networks	63-68
<b>Approaches to explaining AI</b>	Interpretability part 1	15
<b>Architecture and hyperparameters</b>	Neural Networks	52-61
<b>Assessment of the model</b>	Regression Analysis Guide	9, 10
<b>Assumptions of multivariate regression</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	2, 4, 5, 6, 7
<b>Barplots</b>	Data visualization	6
<b>Basic images features</b>	Feature Extraction	4
<b>Bias in Machine Learning</b>	Fairness in Machine Learning	6
<b>Bias Mitigation Strategies</b>	Fairness in Machine Learning	7
<b>Binning</b>	Data Preprocessing	22
<b>Branches and merging</b>	Introduction to GIT	8
<b>Calculating Splits and Growing the Tree</b>	XGBoost	7
<b>Can I trust the Model based on Accuracy?</b>	Interpretability part 1	9,10, 11
<b>Case study</b>	LLM & Ethics	39, 40
<b>Characteristics of the explanation provided by LIME</b>	Interpretability part 4	6
<b>Class weighting</b>	Imbalanced Dataset	9
<b>Classification metrics - Accuracy</b>	Evaluation Metrics	5-7

<b>Classification metrics - Precision, Recall</b>	Evaluation Metrics	8, 9
<b>Clustering</b>	Clustering	6
<b>clustering applications</b>	Clustering	8
<b>clustering in customer analysis</b>	Clustering	9
<b>Clustering in Image Processing</b>	Clustering	11
<b>clustering in medical diagnostics</b>	Clustering	10
<b>Clustering in Text Analysis</b>	Clustering	12
<b>Color features</b>	Feature Extraction	5-7
<b>Combining datasets</b>	Data Preparation (for tabular data)	11-12
<b>Comparison of the Models</b>	Low Code 1	12
<b>Comparison with other Low code /No code tools</b>	No Code	13
<b>Comparison with other methods</b>	Random Forest	9
<b>Connection to a database</b>	Data Preprocessing	6
<b>Core Libraries in Orange</b>	No Code	13
<b>Correlation analysis</b>	Data Preprocessing	18
<b>Create the Model</b>	Low Code 1	13
<b>Creating a multivariate linear regression model</b>	Simple and multivariate linear regression	9
<b>Creating Random Sample</b>	Low Code 1	8
<b>Cross Validation</b>	Time Series Forecasting	9
<b>Data Import and Selection</b>	Data Preprocessing	4
<b>Data Imputation</b>	Data Preparation (for tabular data)	6
<b>Data Loading and Preprocessing</b>	No Code	6
<b>Data normalization</b>	Data Preparation (for tabular data)	13-14
<b>Data Preparation</b>	Data Preprocessing	10
<b>Data Preprocessing Workflow</b>	Data Preprocessing	3
<b>Data Selection and cleaning</b>	Data Preprocessing	9
<b>Data transformation</b>	Data Preprocessing	12
<b>Data type conversion</b>	Data Preprocessing	11
<b>Data types</b>	Data Preparation (for tabular data)	7-9
<b>Data visualization</b>	Data visualization	4
<b>Decision Trees</b>	Random Forest	4
<b>Decision Trees as Base Learners</b>	XGBoost	3
<b>Deep Learning: When to use ML?   When to use DL?</b>	Classification#2_Types_of_Classification_Models	19
<b>Definition of large language models</b>	LLM & Ethics	8
<b>Definition of natural language models</b>	LLM & Ethics	6
<b>Development Environments</b>	Installing Python	6
<b>Differences(Time series forecasting vs Regression)</b>	Time Series Forecasting	8
<b>Different data splits</b>	Data Preprocessing	24

<b>Dimensionality reduction</b>	Feature Extraction	17
<b>Distance measures</b>	Clustering	13-16
<b>Distribution plots - Boxplots</b>	Data visualization	8-9
<b>Drawbacks of lime</b>	Interpretability part 4	13
<b>Dummy Variable Added</b>	Regression Analysis Guide	8
<b>Dummy variables</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	10, 11, 12, 14
<b>Edge detection</b>	Feature Extraction	11, 12
<b>Elements in cv algorithms</b>	Computer Vision	65-74
<b>Embedded algorithms</b>	Feature Engineering	14
<b>Embeddings</b>	LLM & Ethics	16, 17
<b>Ensemble Model</b>	Forecasting Methods	8
<b>Estimation of multivariate linear regression model</b>	Simple and multivariate linear regression	10, 11
<b>Ethical and Practical challenges</b>	Fairness in Machine Learning	8
<b>EU Definition of trustworthy AI</b>	LLM & Ethics	38
<b>Evaluation Interpretability</b>	Interpretability part 1	13
<b>Evaluation metrics</b>	Imbalanced Dataset	6
<b>Evaluation metrics in machine learning</b>	Evaluation Metrics	4
<b>Evaluation the Model</b>	Low Code 2	8
<b>Example of Clasification Models</b>	Low Code 1	6
<b>Exclusive Feature Bundling</b>	LightGBM	7
<b>Exploratory data analysis</b>	Data Preprocessing	16
<b>Extracting statistical features</b>	Feature Extraction	16
<b>Fairness in Machine Learning</b>	Fairness in Machine Learning	4, 5
<b>Feature Extraction from images</b>	Feature Extraction	3
<b>Feature extraction from tabular data</b>	Feature Extraction	15
<b>Feature Extraction from text data</b>	Feature Extraction	18
<b>Feature Extraction using convolution</b>	Feature Extraction	13
<b>Feature Importance</b>	Interpretability part 3	8-10
<b>Feature Selection</b>	Feature Engineering	5,6
<b>Feature Selection types</b>	Feature Engineering	7,8
<b>Fidelity-Interpretability Trade-Off</b>	Interpretability part 4	7
<b>Filter</b>	Computer Vision	21 - 40
<b>Filtering</b>	Feature Extraction	9,10
<b>Finalizing the Model</b>	Low Code 2	9
<b>Fisher Score</b>	Feature Engineering	9,10
<b>Forecast Horizons</b>	Time Series Forecasting	6

<b>Forecast of Y values</b>	Regression Analysis Guide	13
<b>From language to vectors</b>	LLM & Ethics	13
<b>Functionality of Random Forest</b>	Random Forest	6
<b>Generalization</b>	Overfitting	4
<b>Get the Data</b>	Low Code 1	7
<b>GitHub</b>	Introduction to GIT	10
<b>Global machine learning interpretability techniques</b>	Interpretability part 2	8
<b>Global Surrogate Models</b>	Interpretability part 3	11-13
<b>Google Colab</b>	Installing Python	8
<b>Google Collab and GIT</b>	Introduction to GIT	11
<b>Gradient Boosting</b>	Forecasting Methods	9
<b>Gradient Boosting Classifier</b>	Low Code 1	16
<b>Gradient-based One-Side Sampling</b>	LightGBM	9-17
<b>Graph of multivariate linear regression model</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	3
<b>Greedy Bundling</b>	LightGBM	8
<b>Grid Search</b>	Hyperparameter Tuning	6
<b>Hard vs soft clustering</b>	Clustering	7
<b>Heatmaps</b>	Data visualization	10
<b>Histogram-based Algorithm</b>	LightGBM	18
<b>How can LLMs handle textual data/inputs?</b>	LLM & Ethics	12
<b>How to choose your models?</b>	Classification#2_Types_of_Classification Models	22
<b>How to evaluate your classification model?</b>	Classification#3_Evaluating_Classification_Models	5, 6, 7, 8, 9
<b>How to measure Calibration? Reliability Diagram</b>	Classification#3_Evaluating_Classification_Models	16-23
<b>Hyperband Optimization</b>	Hyperparameter Tuning	8
<b>Hyperparameter Tuning</b>	Hyperparameter Tuning	5
<b>Hyperparameters</b>	Hyperparameter Tuning	4
<b>Impact of LLMs</b>	LLM & Ethics	5
<b>Important concepts in computer vision</b>	Computer Vision	20
<b>Important Outputs of setup()</b>	Low Code 1	11
<b>Importing data from a flat file</b>	Data Preprocessing	5
<b>Individual conditional expectation</b>	Interpretability part 2	13,14
<b>Interactions between variables</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	13
<b>Introduction to GIT</b>	Introduction to GIT	5

<b>Introduction to Linear Regression</b>	Simple and multivariate linear regression	2
<b>Jupyter Notebook</b>	Installing Python	7
<b>Key Git Commands</b>	Introduction to GIT	7
<b>Key XGBoost Concepts</b>	XGBoost	8
<b>Legal Issues GDPR</b>	Interpretability part 1	14
<b>LightGBM</b>	Forecasting Methods	10
<b>LIME Image</b>	Interpretability part 4	10,1 1
<b>LIME-Key Ideas</b>	Interpretability part 4	8,9
<b>Linear Models in 1D</b>	Classification#2_Types_of_Classification Models	6
<b>Linear Models in 2D</b>	Classification#2_Types_of_Classification Models	7
<b>Linear Models in 3D</b>	Classification#2_Types_of_Classification Models	8
<b>Linear vs Non-Linear Models</b>	Classification#2_Types_of_Classification Models	21
<b>Lineplots</b>	Data visualization	7
<b>LLM architecture</b>	LLM & Ethics	23
<b>LLM Examples</b>	LLM & Ethics	26-30
<b>LLM Training</b>	LLM & Ethics	24
<b>LLMs and Higher Education</b>	LLM & Ethics	36,37
<b>Loss function</b>	LightGBM	6
<b>Low-Code approach</b>	Low Code 1	2
<b>Machine Learning</b>	Introduction to ML	4
<b>Machine Learning applications becoming pervasive</b>	Interpretability part 1	5
<b>Machine Learning Methods for Time Series</b>	Forecasting Methods	5
<b>Machine Learning Pipeline with PyCaret</b>	Low Code 2	2
<b>Main Component</b>	LightGBM	5
<b>Main findings from aggregated results</b>	LLM & Ethics	31
<b>Merging data tables</b>	Data Preprocessing	7
<b>Minimizing Residuals with Gradient Descent</b>	XGBoost	5
<b>Missing Data</b>	Data Preparation (for tabular data)	4-5
<b>Model Building</b>	No Code	9
<b>Model Evaluation</b>	No Code	10
<b>Modelling preparation</b>	Data Preprocessing	20
<b>Modern CV</b>	Computer Vision	76-85

<b>More advanced techniques of text feature extraction</b>	Feature Extraction	19
<b>Multivariate coefficients of determination</b>	Simple and multivariate linear regression	14
<b>Multivariate regression</b>	Simple and multivariate linear regression	8
<b>Neural Network Basics</b>	Neural Networks	4-28
<b>Neural Network Training</b>	Neural Networks	30-50
<b>No-Code approach</b>	No Code	2
<b>Non-Linear Models</b>	Classification#2_Types_of_Classification Models	9-18
<b>Orange Data Mining</b>	No Code	3
<b>Orange ML: Overview and Features</b>	No Code	4
<b>Orange Workflow</b>	No Code	5
<b>outcomes</b>	Data Preprocessing	25
<b>Overfitting</b>	Overfitting	5-8
<b>Oversampling the minority class</b>	Imbalanced Dataset	7
<b>Overview of XGBoost and Boosting</b>	XGBoost	2
<b>Padding</b>	Computer Vision	42-48
<b>Partial dependency plot</b>	Interpretability part 2	9-12
<b>Plot structure</b>	Data visualization	11
<b>Plotting the Model</b>	Low Code 2	4, 5, 6, 7
<b>Pooling</b>	Computer Vision	58-63
<b>Predicting with the Model</b>	Low Code 2	10
<b>Probabilistic Classification</b>	Classification#3_Evaluating_Classification_Models	10, 11, 12, 13
<b>Probabilistic decision tree</b>	Classification#3_Evaluating_Classification_Models	15
<b>Probabilistic Linear Classifier</b>	Classification#3_Evaluating_Classification_Models	14
<b>Proper ML Pipeline</b>	Classification#1_What_is_Classification	24, 25
<b>Pull Request</b>	Introduction to GIT	9
<b>PyCaret</b>	Low Code 1	3, 4, 5
<b>Python modules</b>	Feature Extraction	14
<b>Quantity sold = f(advertising costs)</b>	Regression Analysis Guide	11
<b>Random Forest</b>	Random Forest	5
<b>Random Forest Classifier</b>	Low Code 1	14
<b>Random Search</b>	Hyperparameter Tuning	7
<b>Regression</b>	Time Series Forecasting	7

<b>Regression - Least Squares Method</b>	Simple and multivariate linear regression	5, 6, 7
<b>Regression metrics</b>	Evaluation Metrics	10
<b>Regression metrics - Mean Absolute Error (MAE)</b>	Evaluation Metrics	11
<b>Regression metrics - Mean Squared Error (MSE)</b>	Evaluation Metrics	12
<b>Regression metrics - Root Mean Squared Error (RMSE)</b>	Evaluation Metrics	13
<b>Regularization, Overfitting, and Learning Rate</b>	XGBoost	4
<b>Reinforcement Learning</b>	Introduction to ML	5,12-13
<b>Relation of NLP, foundation model, and LLM</b>	LLM & Ethics	9
<b>Removing unseen data</b>	Low Code 1	9
<b>Repositories - local and remote</b>	Introduction to GIT	6
<b>Residuum</b>	Regression Analysis Guide	12
<b>Ridge Classifier</b>	Low Code 1	15
<b>Sampling</b>	Data Preprocessing	21
<b>Save/Load Model for Production</b>	Low Code 2	11
<b>Scatterplots</b>	Data visualization	5
<b>Semi-Supervised Learning</b>	Introduction to ML	5
<b>Sequential forward selection</b>	Feature Engineering	11,12
<b>Setting up the PyCaret environment Now</b>	Low Code 1	10
<b>Signal decomposition</b>	Data Preprocessing	17
<b>Significance of individual variables</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	8
<b>Significance of the whole model</b>	Advanced Concepts in Linear Regression: Assumptions, Dummy Variables, and Interactions	9
<b>Some Properties of Interpretations</b>	Interpretability part 1	12
<b>Splitting data for model training and evaluation</b>	Data Preprocessing	23
<b>Starting with a Single Leaf Node</b>	XGBoost	6
<b>Steps of regression analysis</b>	Regression Analysis Guide	2, 3, 4, 5, 6, 7
<b>Strengths and Weaknesses of Random Forest</b>	Random Forest	8
<b>Stride</b>	Computer Vision	50-56

<b>Submodular selection (SP-LIME)- Key Ideas</b>	Interpretability part 4	12
<b>Subplots</b>	Data visualization	12
<b>Supervised Classification</b>	Classification#1 What is Classification	6
<b>Supervised Learning</b>	Introduction to ML	5-10,13
<b>Supervised ML Pipeline</b>	Classification#1_What_is_Classification	13-23
<b>Supervised Regression</b>	Classification#1 What is Classification	7, 8
<b>Support Vector Machine</b>	Forecasting Methods	6
<b>Support Vector Regression</b>	Forecasting Methods	7
<b>Taxonomy of Interpretability methods</b>	Interpretability part 2	7
<b>Texture features</b>	Feature Extraction	8
<b>The meaning of simple linear regression</b>	Simple and multivariate linear regression	3
<b>Time Series</b>	Time Series Forecasting	4
<b>Time Series Cross Validation</b>	Time Series Forecasting	10
<b>Time Series pattern</b>	Time Series Forecasting	5
<b>Tokenize language</b>	LLM & Ethics	14,15
<b>Trade off between Accuracy and Interpretability</b>	Interpretability part 2	6
<b>Transformation</b>	No Code	7
<b>Tuning the Model</b>	Low Code 2	3
<b>Types of Deep Learning Architectures</b>	Classification#2_Types_of_Classification Models	20
<b>Types of joins</b>	Data Preprocessing	8
<b>Types of Machine Learning Tasks</b>	Classification#1_What_is_Classification	4, 5, 9, 11
<b>Uncertainty Quantification</b>	Time Series Forecasting	11
<b>Underfitting</b>	Overfitting	9-10
<b>Undersampling the majority class</b>	Imbalanced Dataset	8
<b>Unsupervised</b>	No Code	11
<b>Unsupervised Learning</b>	Introduction to ML	5,11,13
<b>Use of Python</b>	Installing Python	5
<b>Visualization</b>	No Code	8
<b>What do machines see?</b>	Computer Vision	10-18
<b>What does it look like to interact with LLMs?</b>	LLM & Ethics	10
<b>What is a foundation model?</b>	LLM & Ethics	7
<b>What is an imbalanced Dataset?</b>	Imbalanced Dataset	4

<b>What is Computer Vision?</b>	Computer Vision	5, 6, 7
<b>What is interpretability?</b>	Interpretability part 1	4
<b>What is ML Model?</b>	Classification#2_Types_of_Classification Models	5
<b>What is version control?</b>	Introduction to GIT	4
<b>Why explainability</b>	Interpretability part 1	6
<b>Why Explainability: Verify the ML Model / System</b>	Interpretability part 1	8
<b>Why is an imbalanced Dataset a Problem?</b>	Imbalanced Dataset	5
<b>Why is the model making these predictions?</b>	Interpretability part 1	7
<b>Why Python?</b>	Installing Python	4
<b>Why we need data preprocessing?</b>	Data Preprocessing	2
<b>Why we need version control?</b>	Introduction to GIT	3
<b>Wrapper algorithms</b>	Feature Engineering	13
<b>XGBoost Efficiency Features</b>	XGBoost	9
<b>XGBoost in Classification Problems</b>	XGBoost	11
<b>XGBoost in Regression Problems</b>	XGBoost	10
	Classification#1_What_is_Classification	6,7,8
	Classification#1_What_is_Classification	10
	Classification#1_What_is_Classification	12
	Data preprocessing	19
	Introduction to ML	5, 10
	Interpretability part 2	5