



AdvancedAnalytics
.Academy

www.advancedanalytics.academy

Introduction to Advanced Analytics

Stefan Weingaertner
Stuttgart, 11/12/2017



Agenda

1. Introduction to Advanced Analytics
2. Process Models
3. Myths and Pitfalls
4. Categorization of Advanced Analytics Algorithms
5. Fields of Application (exemplary use cases)
6. Challenges
7. Roles in a Data Science Team
8. The Value of Analytics

1.

Introduction to Advanced Analytics

3

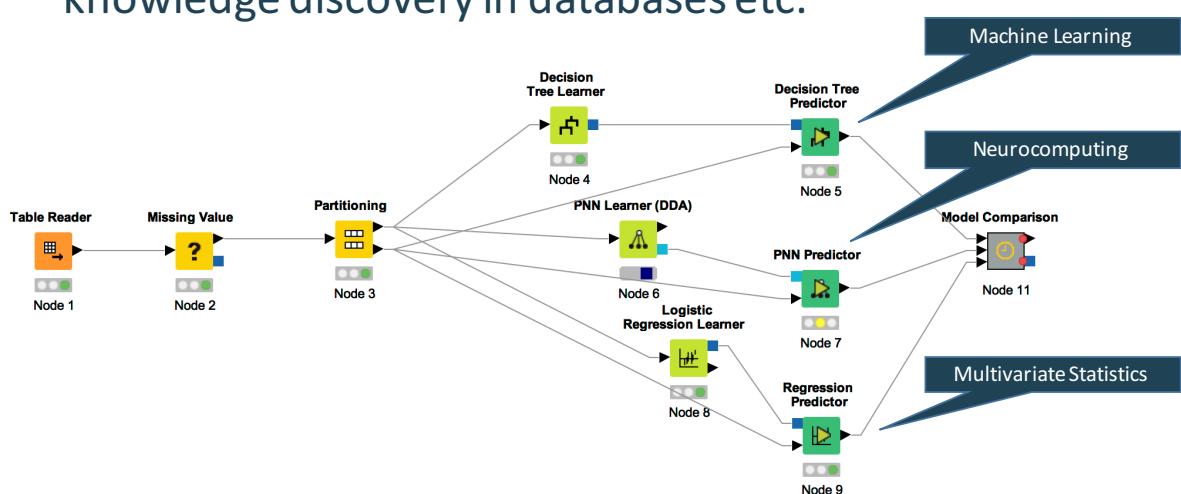
This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



1. What is Advanced Analytics?

- ...is a **combination** of different academic fields like multivariate statistics, artificial intelligence, machine learning, pattern recognition, neurocomputing, knowledge discovery in databases etc.



4

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

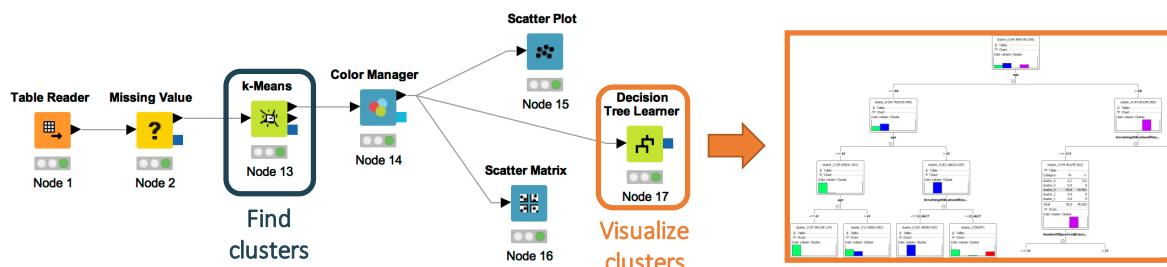
© AdvancedAnalytics.Academy GmbH



1. What is Advanced Analytics?

- ...is the **process** of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cut costs, or identify business critical hidden patterns.

Combination of Cluster Analysis & Decision Trees



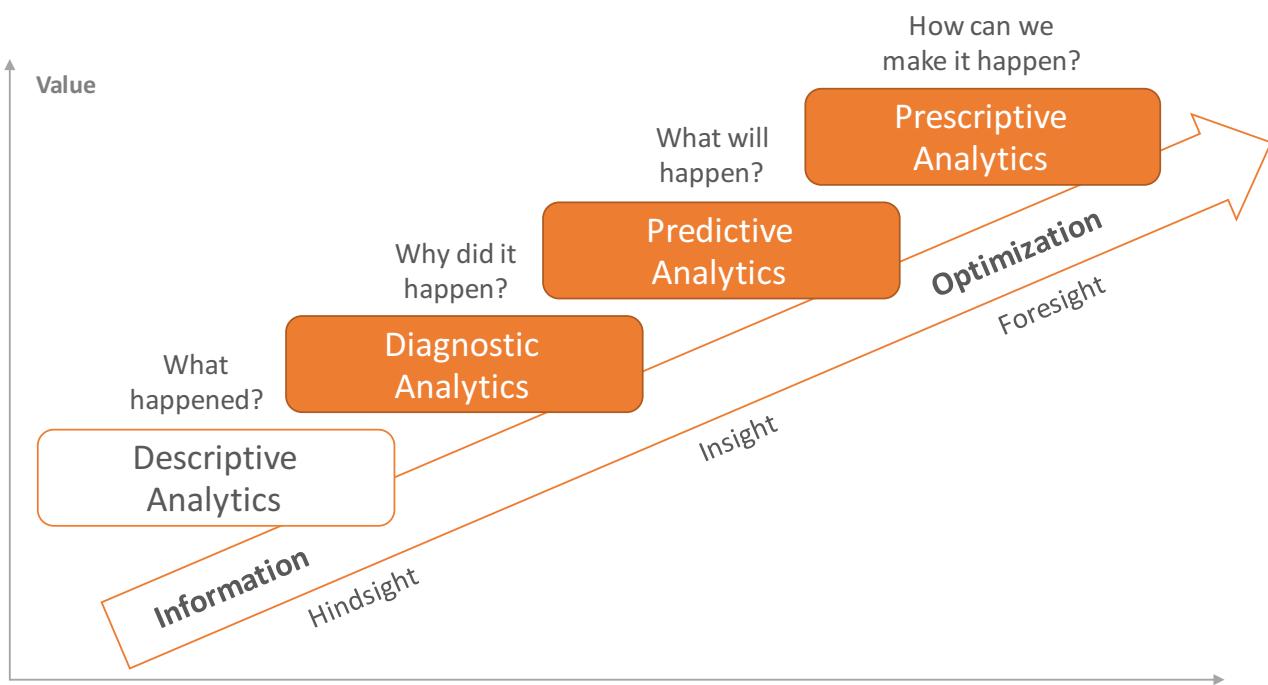
5

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Advanced Analytics – Beyond Descriptive Analytics



Source: Gartner Group, 2013

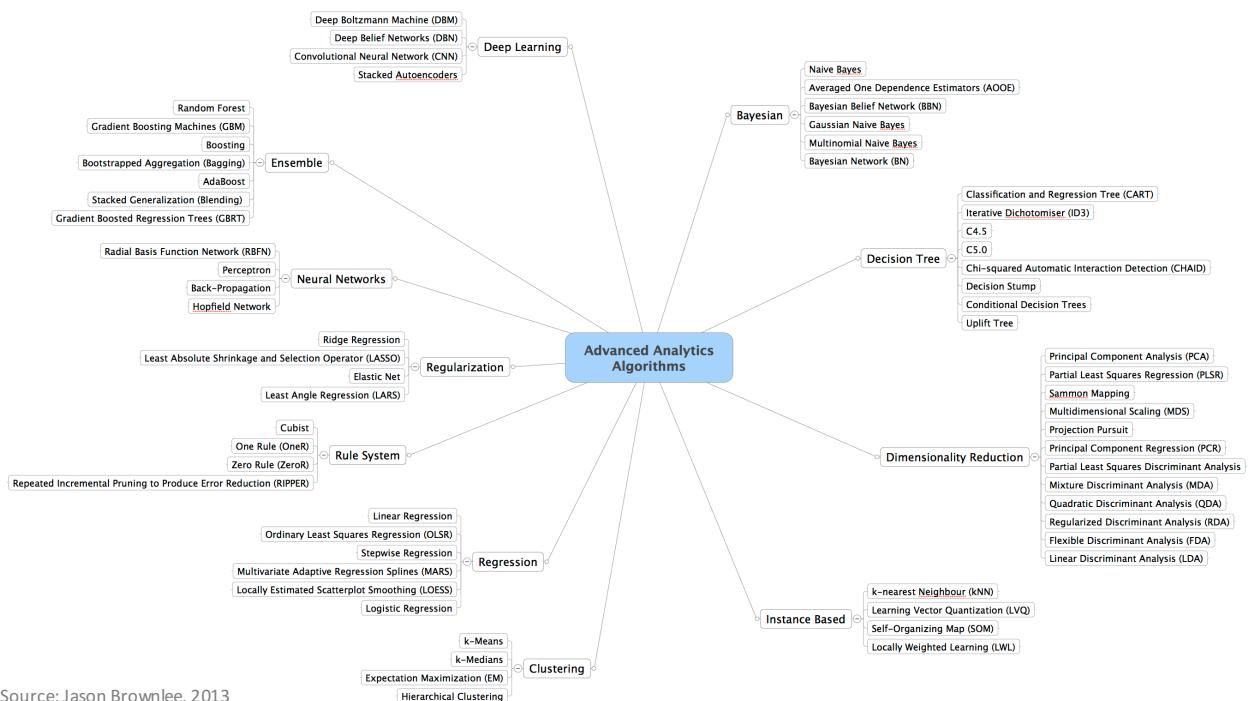
6

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



An overview of Advanced Analytics Algorithms



Source: Jason Brownlee, 2013

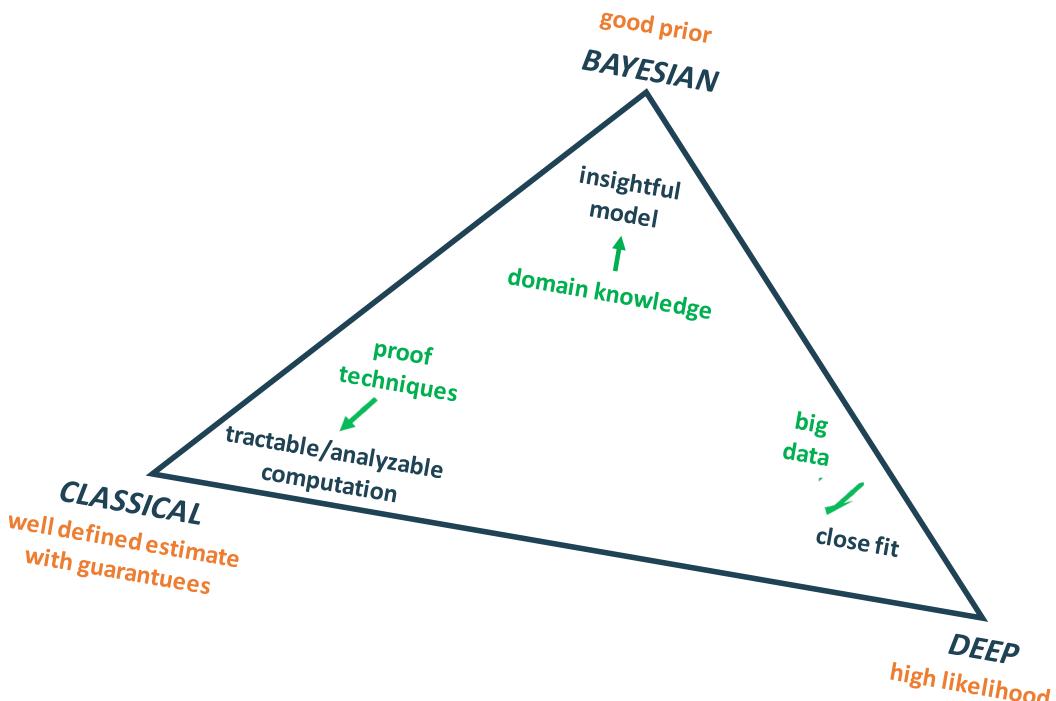
7

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



The Three Cultures of Machine Learning



Source: Jason Eisner, 2015

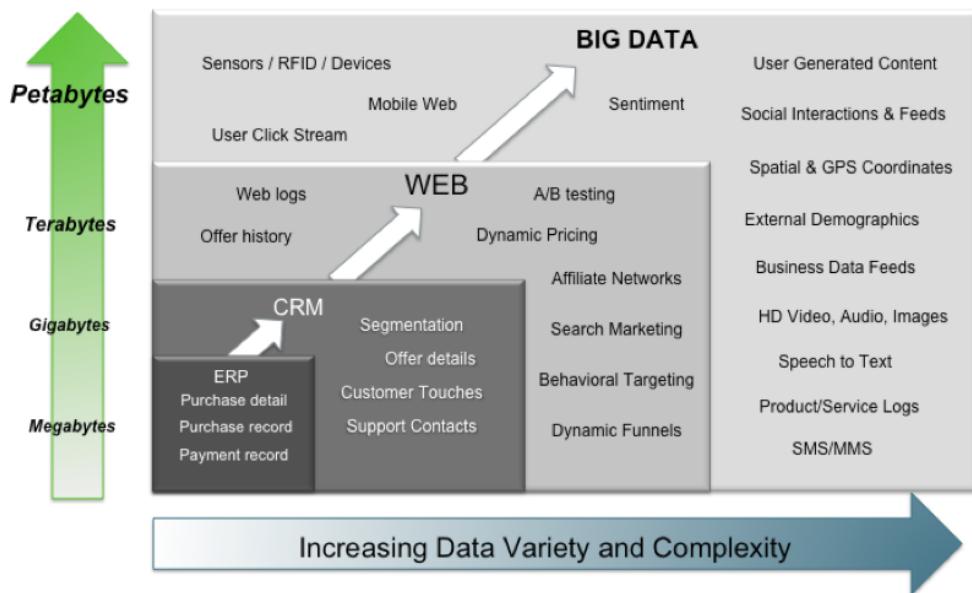
8

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Determining factors for the rise of Big Data Analytics



Source: Contents of above graphic created in partnership with Teradata, Inc.

9

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Developments

- The rise of open source...
- Algorithm development, e.g.
 - Ensemble Models
 - Deep Learner
- Hyper Parameterization
- Parallel Processing
- Online Learning
- Analytics Automation

10

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



2. Process Models

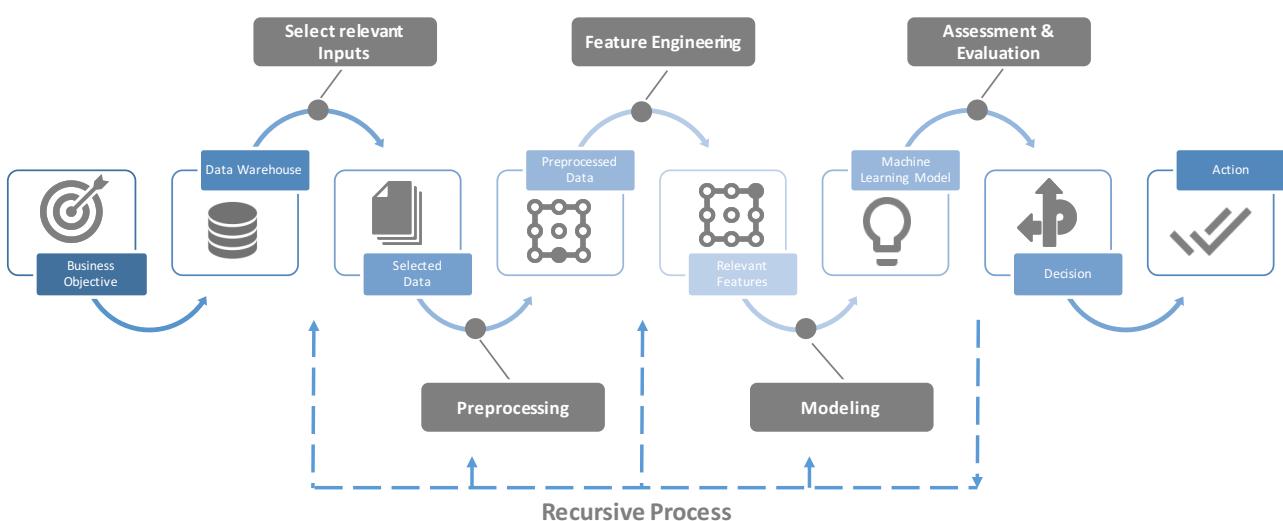
11

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



2. A typical Advanced Analytics Process



12

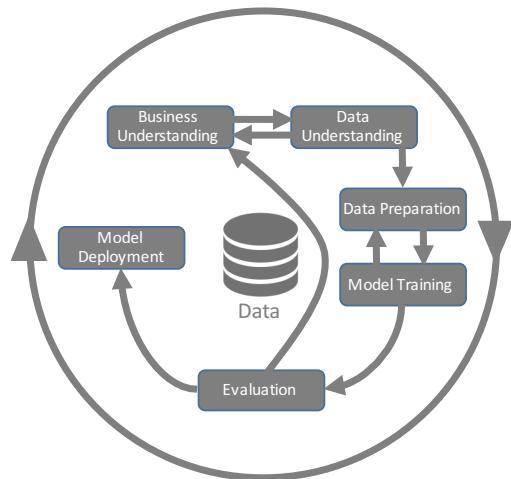
This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



2. CRISP-DM

- CRISP-DM = Cross Industry Standard Process for Data Mining.
- A data mining process model that describes commonly used approaches that expert data miners use to tackle problems.
- CRISP-DM breaks the process of data mining into six phases:
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Model Training
 - Evaluation
 - Model Application



13

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



3. Advanced Analytics Myths and Pitfalls

14

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH





“

If you've got terabytes of data, and you're relying on data mining to find interesting things in there for you, you've lost before you've even begun.

“

Herb Edelstein



3. Advanced Analytics Myths

- Advanced Analytics provides instant crystal ball-predictions
- Advanced Analytics is all about algorithms
- Advanced Analytics is all about predictive accuracy
- A big data warehouse is a requirement for Advanced Analytics
- Advanced Analytics is all about vast quantities of data
- Advanced Analytics is for large companies with lots of customer data
- Only those who know every aspect of the technology must perform Advanced Analytics



3. Advanced Analytics Pitfalls

- Buried under mountains of data
- The Mysterious Disappearing Terabyte
- Disorganized Advanced Analytics
- Insufficient business knowledge
- Insufficient data knowledge
- Erroneous assumptions, courtesy of the experts
- Incompatibility of Advanced Analytics tools

17

This educational material was produced for the Machine Learning Workcamp at
IHK Region Stuttgart December 2017. The copyright is with
AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



4. Categorization of Advanced Analytics Algorithms

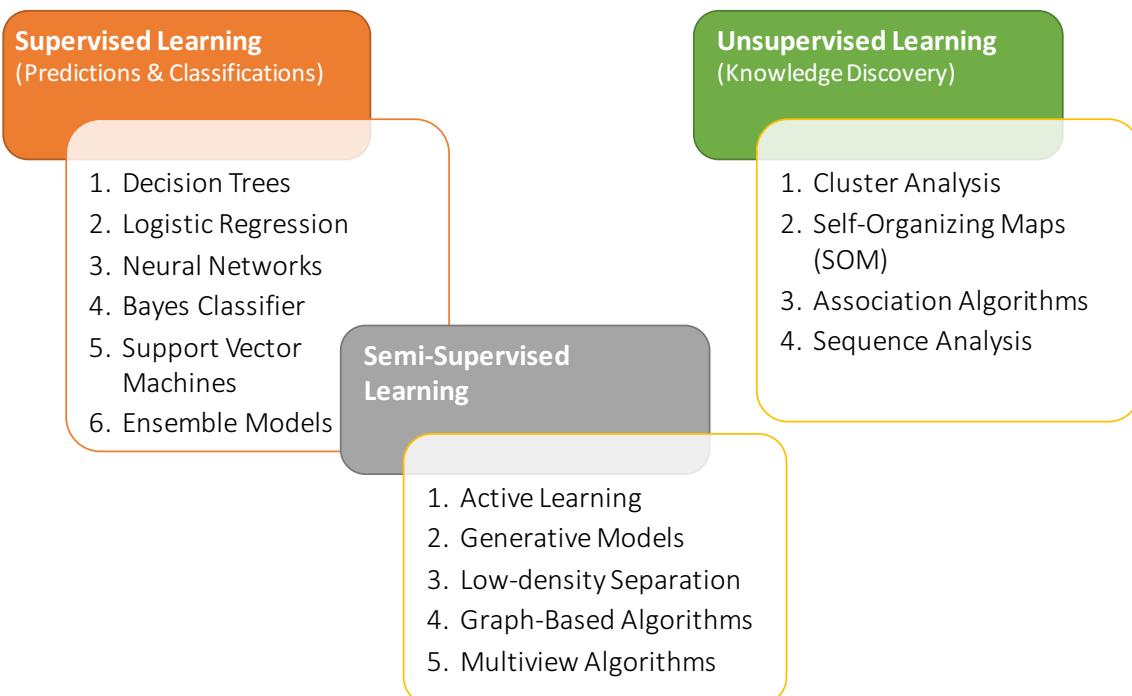
18

This educational material was produced for the Machine Learning Workcamp at
IHK Region Stuttgart December 2017. The copyright is with
AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



4. Categorization of Advanced Analytics Algorithms



19

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



5. Fields of Application (exemplary Use Cases)

20

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



5. Fields of Application (examples)

Customer Segmentation

Customer segmentation, also referred to as market segmentation, is the process of finding **homogenous** sub-groups within a **heterogeneous** aggregate customer base.

Propensity Modeling (Churn, Next Best Offer etc)

1. Predict customer churn by assessing their propensity of risk to churn.
2. Predict customer need and behavior by assessing their propensity to buy a product.

Association Analysis (eg. Market Basket Analysis)

Association rules are employed today in many application areas including market basket analysis, web usage mining, intrusion detection and bioinformatics.

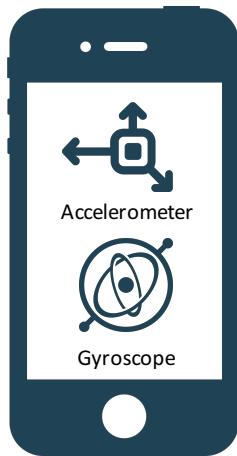
Fraud Detection & Money Laundering

Anticipate illegal or suspicious activities and transactions – such as identity theft, insurance fraud and money laundering by applying predictive analytics methods.

IoT Analytics Use Case Activity Detection with DSP & Machine Learning

IoT Analytics

Overview



Smart Phone



Digital Signal Processing & Machine Learning



Activity Classification

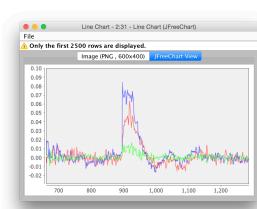
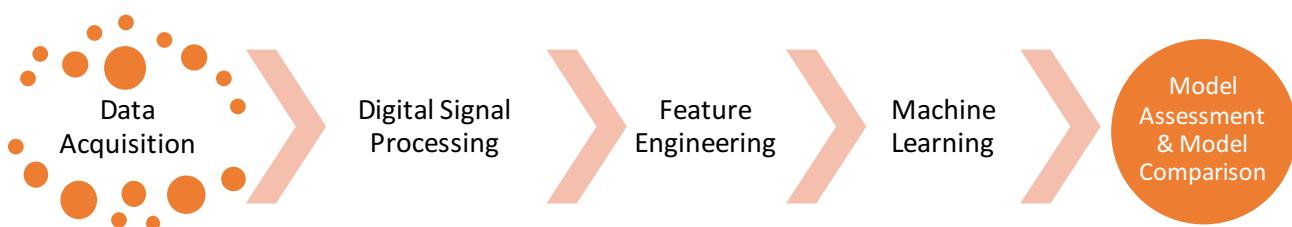
23

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Analytical Process



Confusion Matrix - GBD-Scorer (Model)						
	WALKING	WALKING_UPSTAIRS	WALKING_DOWNSTAIRS	SITTING	STANDING	LAYING
WALKING	445	14	0	0	0	0
WALKING_UPSTAIRS	68	593	20	0	0	0
WALKING_DOWNSTAIRS	0	0	279	0	0	0
SITTING	0	0	0	216	235	0
STANDING	0	0	5	521	0	0
LAYING	0	0	0	0	0	537

24

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



6.

Challenges in Advanced Analytics

25

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



AdvancedAnalytics
.Academy

6. Challenges in Advanced Analytics

- **Prediction Type**
 - Which prediction type should be applied?
(classifications, rankings, estimates)
- **Input Reduction**
 - Which useful input variables can be used to predict the target?
- **Optimization of model complexity**
 - Keep it as simple as possible and as complex as necessary!
- **Generalization of models**
 - How do models behave on new data?
- **Optimization of input data volume**
 - Apply intelligent data sampling!

26

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



AdvancedAnalytics
.Academy

6. Prediction Types for Predictive Modeling

Training Data

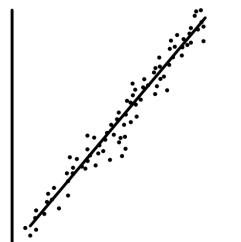
Observations

categorical or numeric input and target measurements

	INPUT VARIABLES				TARGET
	■	■	■	■	■
	■	■	■	■	■
	■	■	■	■	■
	■	■	■	■	■
	■	■	■	■	■

Predictive Model

a formal representation of the input and target association



Icon by Robbe de Clerck (Noun Project)

27

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



6. Prediction Types for Predictive Modeling

INPUT VARIABLES

■	■	■	■
■	■	■	■
■	■	■	■
■	■	■	■
■	■	■	■
■	■	■	■
■	■	■	■

PREDICTION

■
■
■
■
■
■
■

Predictions

output of the predictive model given a set of input measurements

28

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



6. Prediction Types for Predictive Modeling

INPUT VARIABLES				PREDICTION
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■
■	■	■	■	■

Classifications
Rankings
Estimates

Classifications
Rankings
Estimates

Customer cancels contract (YES or NO)
Customer A has a higher propensity to churn than customer B
Customer A has a churn-probability of 75%.

29

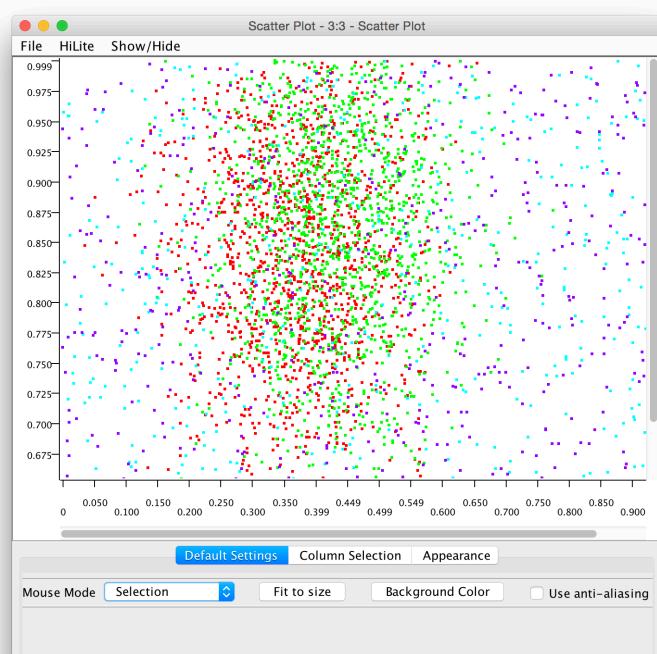
This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



6. Dimensionality Reduction

- Identify and reject redundant and irrelevant input variables



30

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

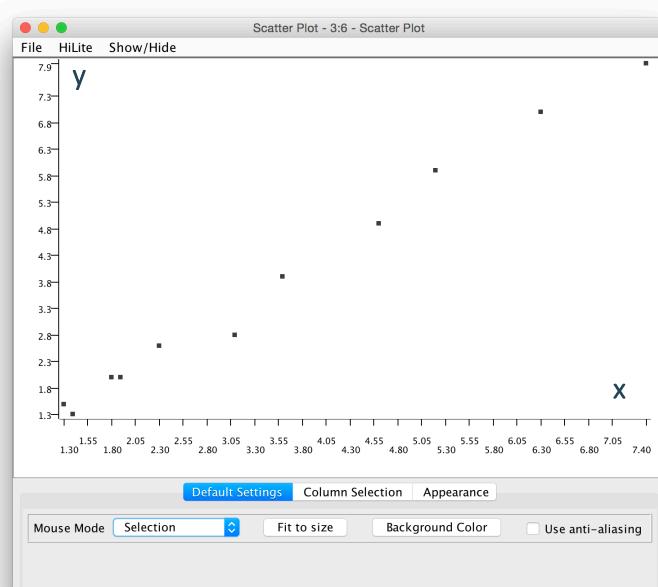
© AdvancedAnalytics.Academy GmbH



6. Dimensionality Reduction

- **Redundancy**

Variable y contains the same information as variable x.



31

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

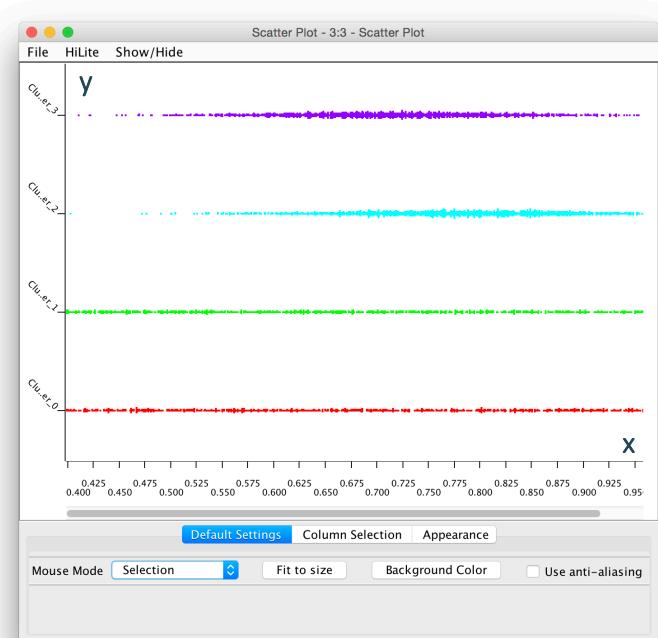
© AdvancedAnalytics.Academy GmbH



6. Dimensionality Reduction

- **Irrelevancy**

Predictions change with variable y but not with Variable x.



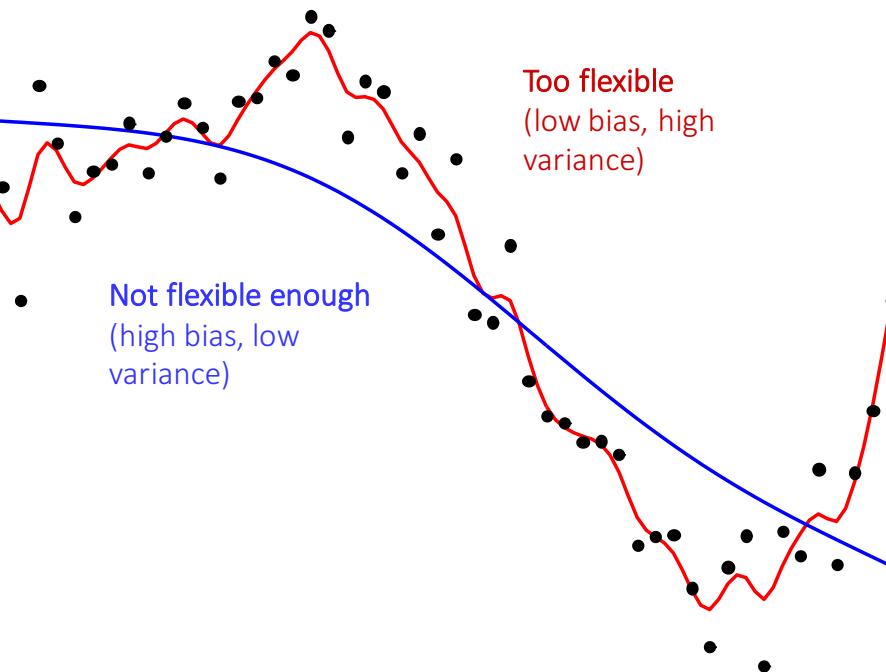
32

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



6. Optimization of model complexity



33

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

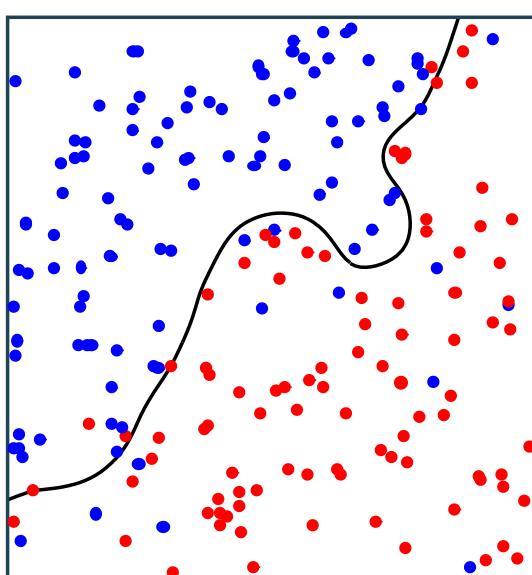
© AdvancedAnalytics.Academy GmbH



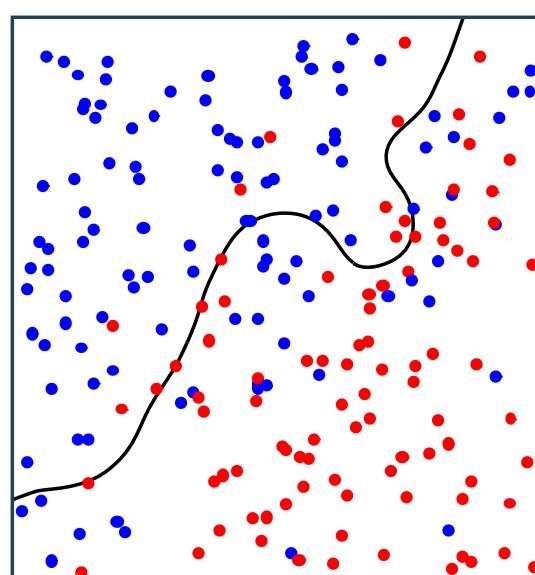
6. Optimization of model complexity

Overfitting (complex model)

TRAINING DATA



TEST DATA



34

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

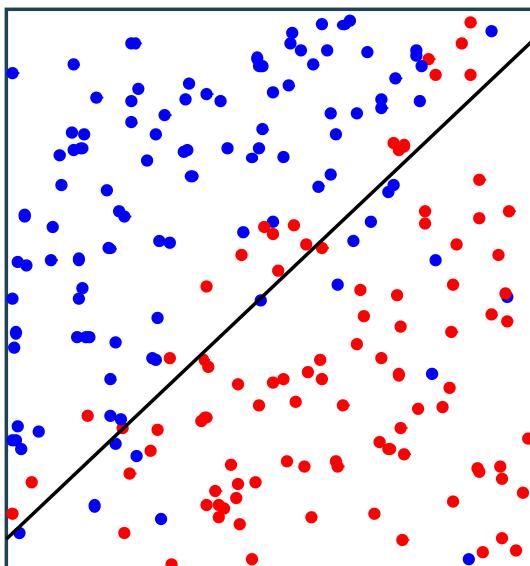
© AdvancedAnalytics.Academy GmbH



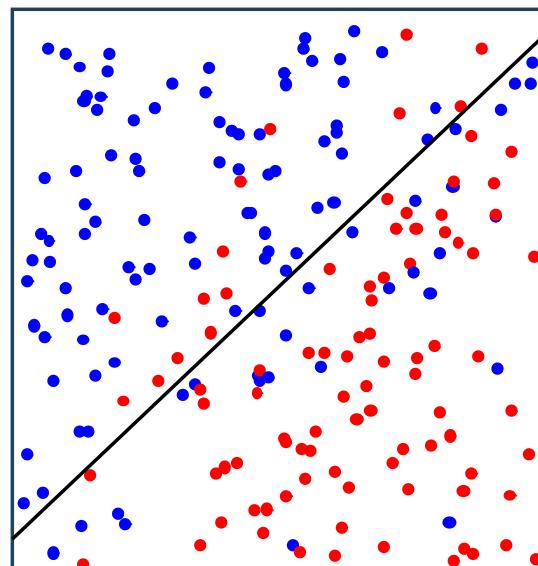
6. Optimization of model complexity

Better fit with simpler model approach

TRAINING DATA



TEST DATA



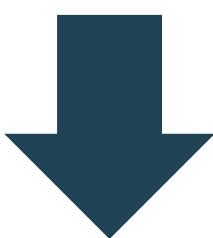
This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



6. Optimization of model complexity

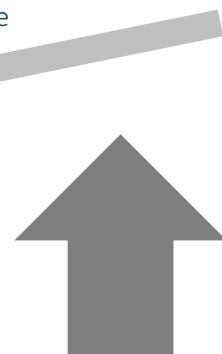
UNDERFITTING

- Reduced prediction quality due to lacking consideration of (non-linear) associations
- High misclassification rate (low accuracy)
- High bias, low variance



OVERFITTING

- Models based on accidentally characteristics
- Bad generalization of models
- Low bias, high variance



! Keep models as simple as possible and as complex as necessary !

6. Generalization of models

• Challenge

- Build high-performance predictive models that generalize well to new data!
- Extend the half-life validity of predictive models even if they are applied to unknown data!

• Out-of-sample

- *Constraint:* the model's input training data set is based on a sample (random or stratified sample).
- Being "out-of-sample-proofed" the model shows the same predictive power (assessment quality values) as being trained on the complete basic population.

• Out-of-time

- *Constraint:* the model's input data set contains observations of a specific time window (e.g. observations of the last 18 months).
- Being "out-of-time-proofed" the model shows the same predictive power (assessment quality values) as being applied to future observations.

37

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

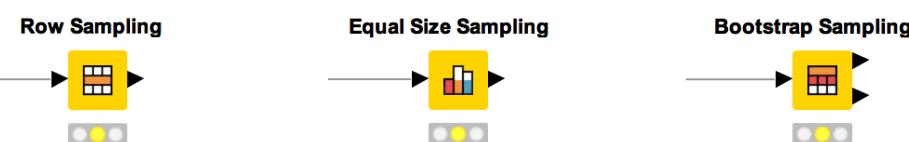
© AdvancedAnalytics.Academy GmbH



6. Optimization of input data size

• Data sampling

- Random sampling eliminates bias by giving all individuals an equal chance to be chosen
- If drawn representatively a model based on a sample shows the same predictive power as being trained on the complete basic population.
- Sampling always implies cost and performance savings



38

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

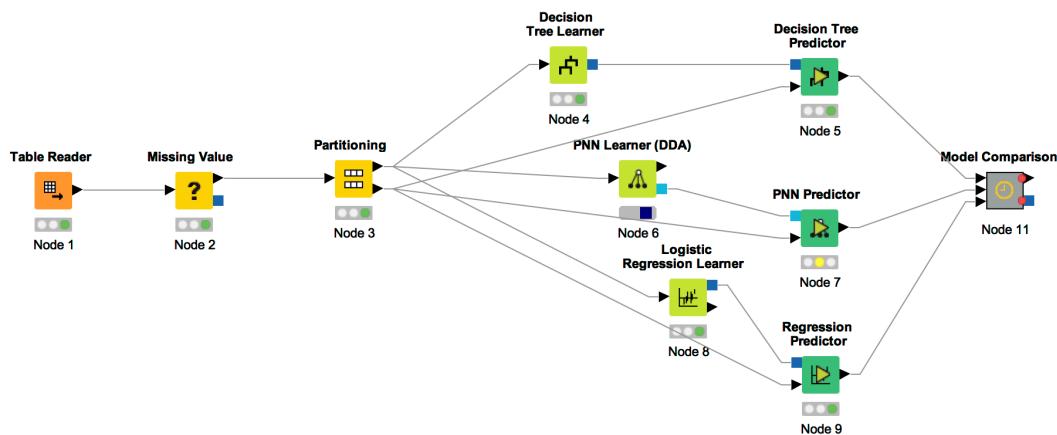
© AdvancedAnalytics.Academy GmbH



6. Optimization of input data size

• Data Partition

- Partitioning of input data into training, validation and test data.
- Training data is used for **model training**.
- Validation and Test data is used for selection of **best model topology** and for **model testing**.



39

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



7.

Roles in a Data Science Team

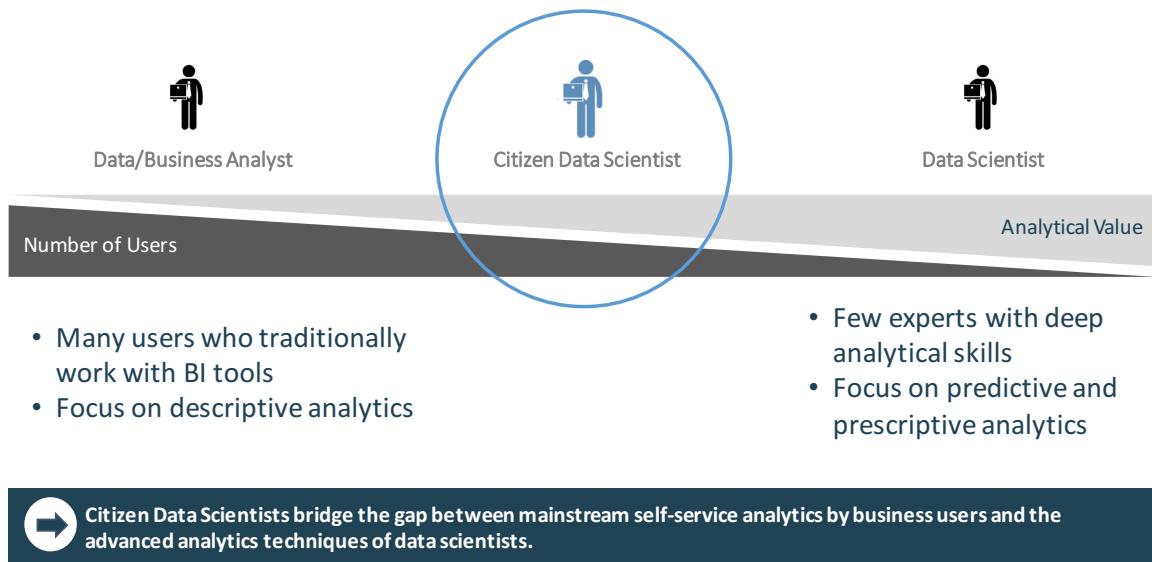
40

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



The role of the Citizen Data Scientist (Gartner, 2017)



41

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



The role of the Citizen Data Scientist (Gartner, 2017)

- A Citizen Data Scientist is a person who creates or generates models that use advanced diagnostic analytics or predictive and prescriptive capabilities, but whose primary job function is outside the field of statistics and analytics.
- Citizen Data Scientists bridge the gap between mainstream self-service analytics by business users and the advanced analytics techniques of data scientists.
- They are able to perform sophisticated analysis that would previously have required more expertise, enabling them to deliver advanced analytics without having the skills that characterize data scientists.

42

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



8.

The Value of Analytics

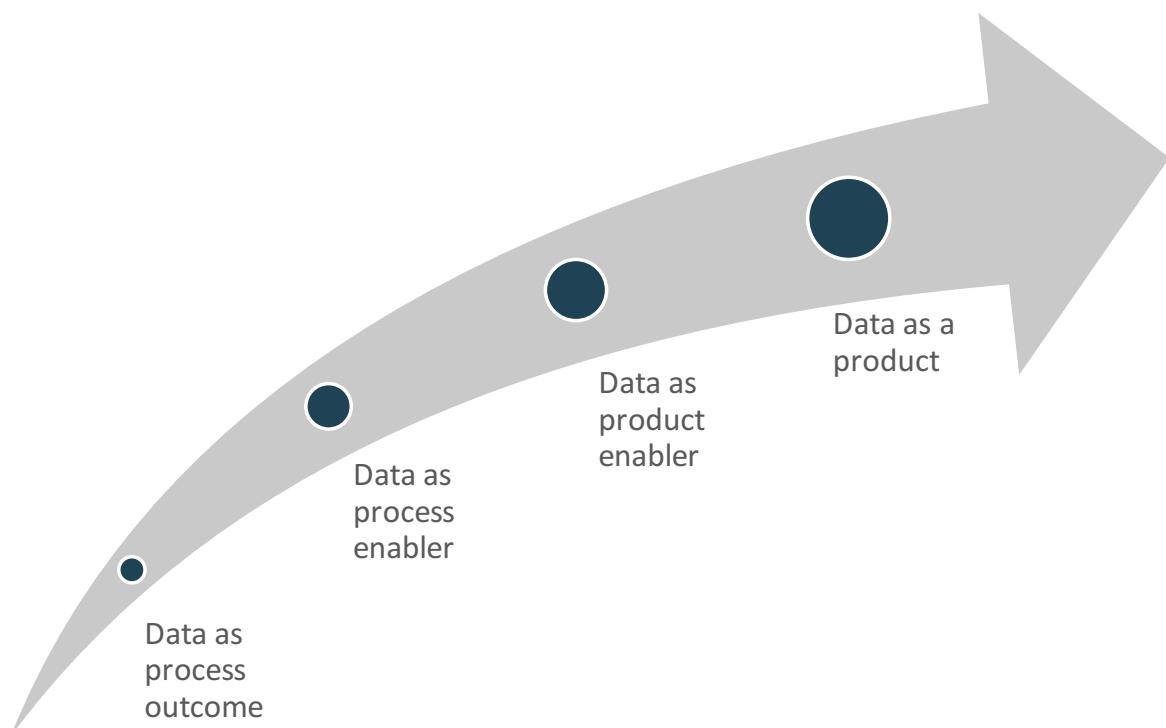
43

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



8. Different roles of data



44

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



8. Data as...

- **process outcome**
 - Measurements, KPI
 - Quality assurance
 - Support main processes
- **process enabler**
 - virtual data
 - used to control MES, ERP, PP
 - E-commerce, electronic cash
- **product enabler**
 - IoT, e.g. smart home, smart wearables, low-cost health/medical equipment
 - Sharing economy, e.g. ride-sharing, esp. mobile/hyper local
- **the product itself**
 - market research data, data broker
 - GPS navigation systems
 - "iTunes/Netflix"



Depending on the overall data strategy the impact and setup of advanced analytics initiatives are completely different.

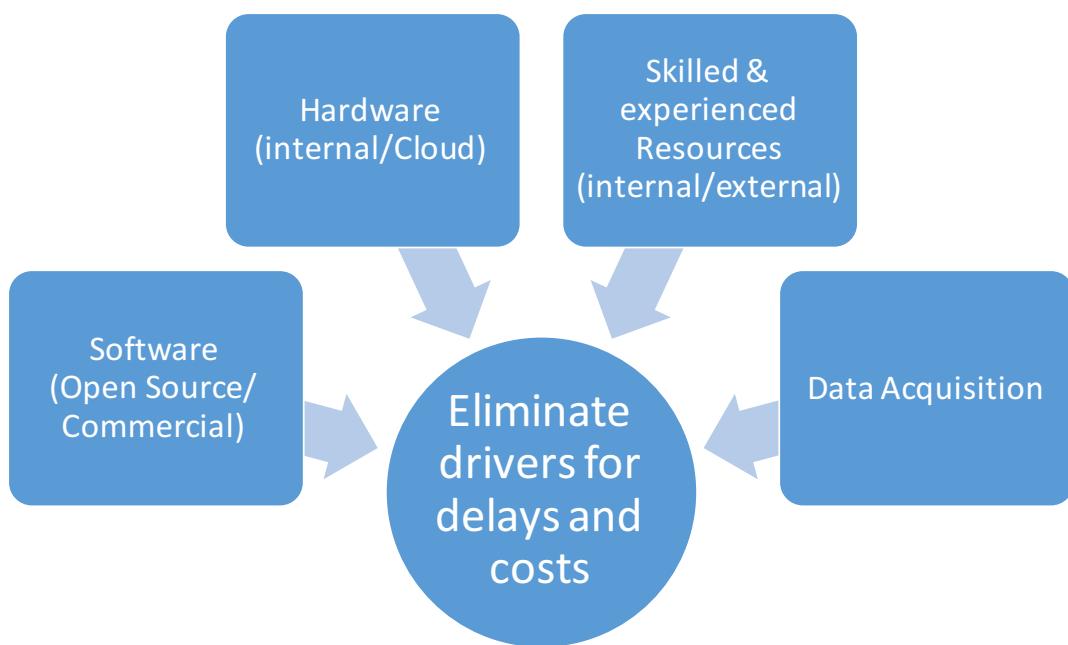
45

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Best practice: Establish an analytics playground for a permanent test&learn environment



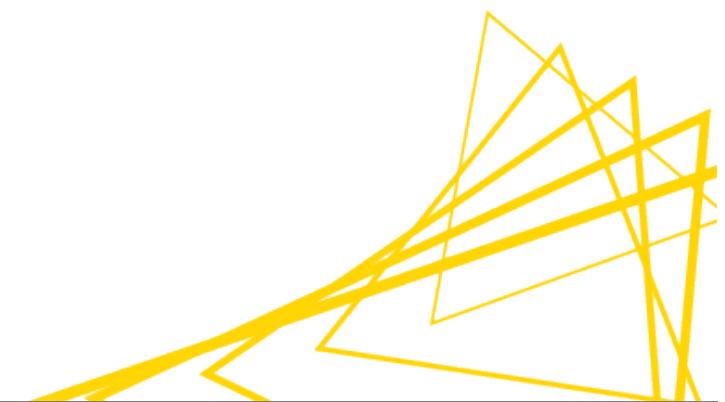
46

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

© AdvancedAnalytics.Academy GmbH



Overview KNIME Analytics Platform



Copyright © 2017 KNIME.com AG

47



Additional Resources

KNIME pages (www.knime.org)

- **SOLUTIONS** for example workflows
- RESOURCES/**LEARNING HUB** www.knime.org/learning-hub
- RESOURCES/**NODE GUIDE** <https://www.knime.org/nodeguide>

KNIME Tech pages (tech.knime.org)

- **FORUM** for questions and answers
- **DOCUMENTATION** for docs, FAQ, changelogs, ...
- **COMMUNITY CONTRIBUTIONS** for dev instructions and third party nodes

KNIME TV on YouTube <https://www.youtube.com/user/KNIMETV>

What is KNIME Analytics Platform?

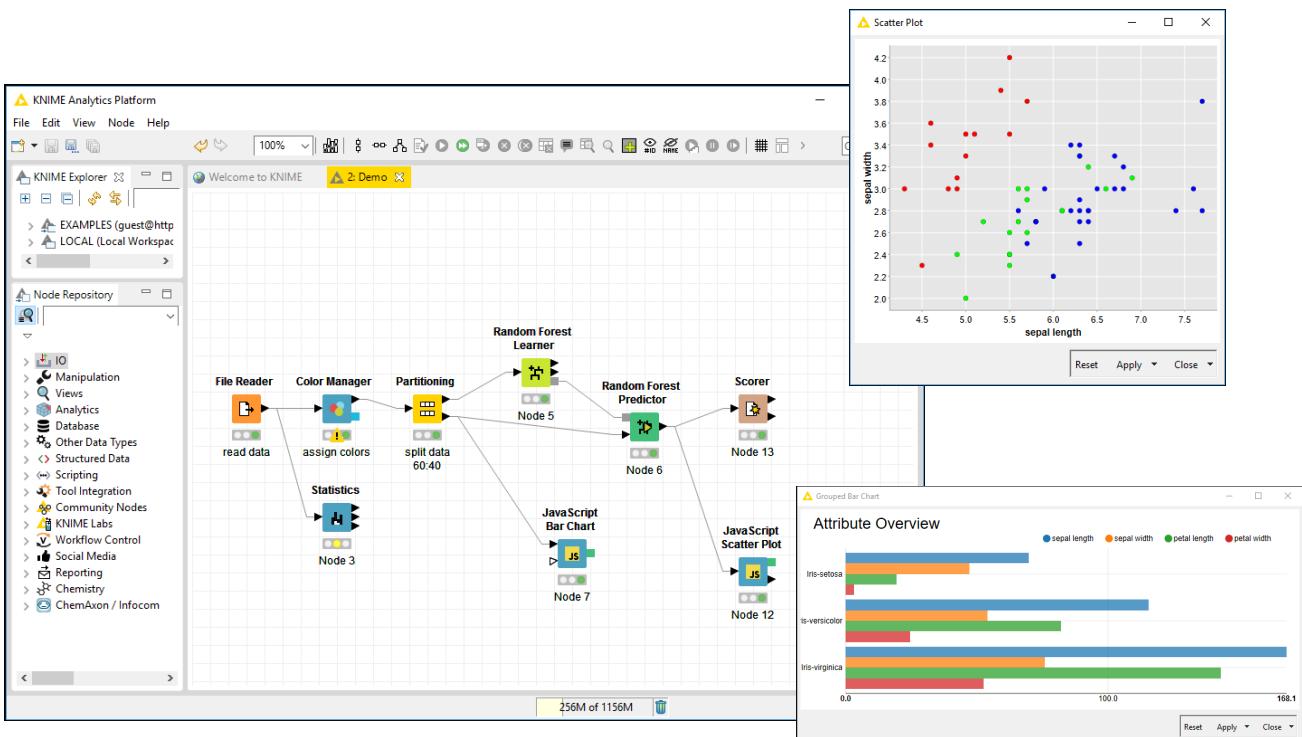
- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
 - Text Mining
 - Network Mining
 - Cheminformatics
 - Weka machine learning
 - Many integrations, such as Java, R, Python, etc.

49

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



The KNIME® Analytics Platform



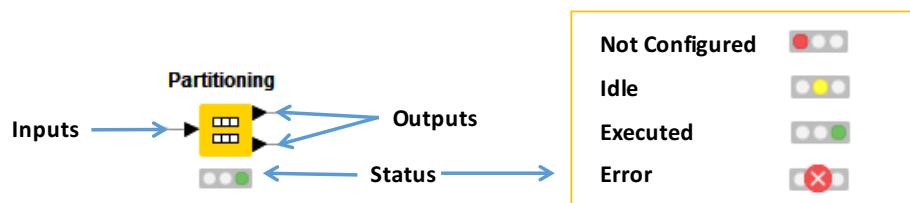
50

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

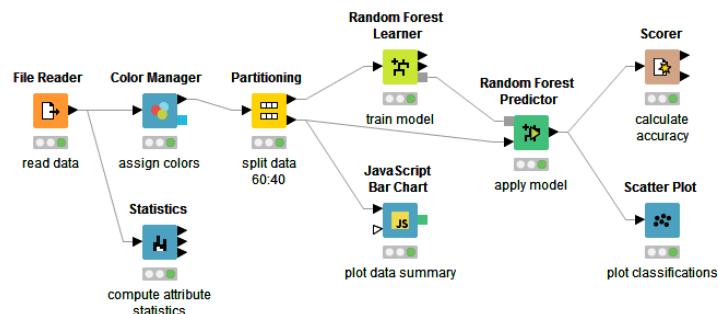


Visual KNIME Workflows

NODES perform tasks on data



Nodes are combined to create
WORKFLOWS

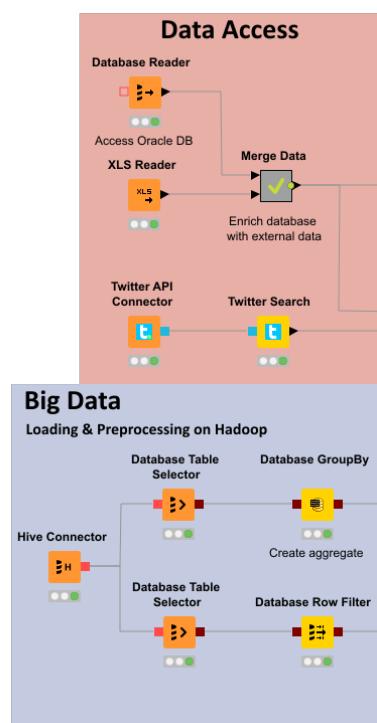


51

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Data Access



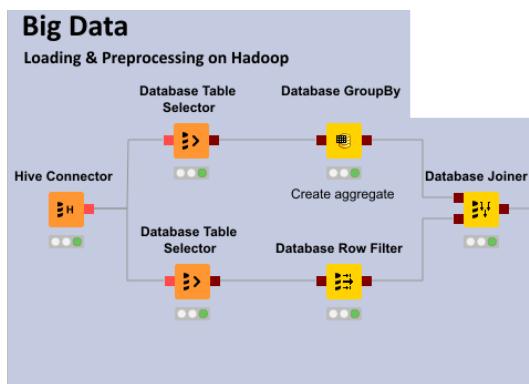
- **Databases**
 - MySQL, PostgreSQL
 - any JDBC (Oracle, DB2, MS SQL Server)
- **Files**
 - Csv, txt
 - Excel, Word, PDF
 - SAS, SPSS
 - XML
 - PMML
 - Images, texts, networks, chem
- **Web, Cloud**
 - REST, Web services
 - Twitter, Google

52

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Big Data



- Spark
- HDFS support
- Hive
- Impala
- HP Vertica
- In-database processing

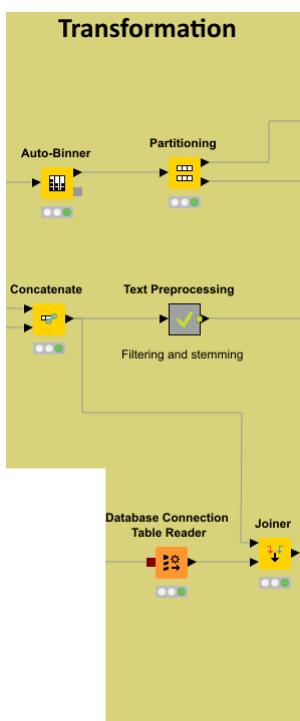


53

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

 AdvancedAnalytics
.Academy

Transformation



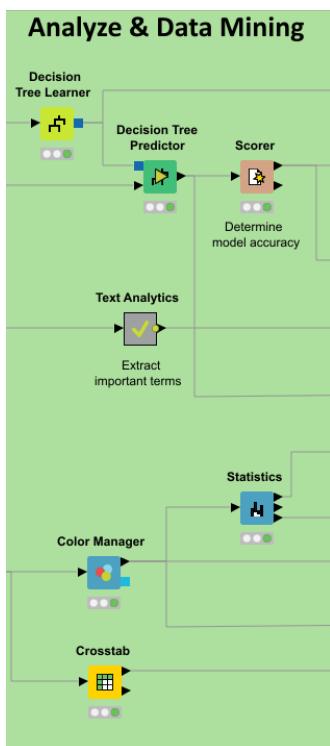
- Preprocessing
 - Row, column, matrix based
- Data blending
 - Join, concatenate, append
- Aggregation
 - Grouping, pivoting, binning
- Feature Creation and Selection

54

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.

 AdvancedAnalytics
.Academy

Analyze & Data Mining



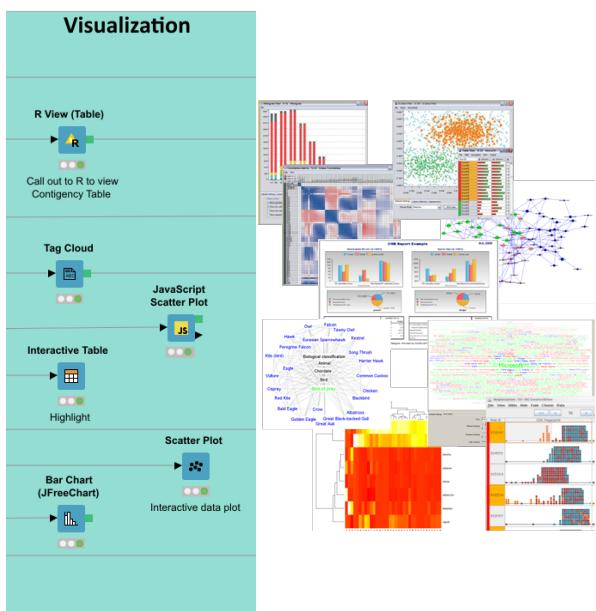
- Regression
 - Linear, logistic
- Classification
 - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
 - k-means, DBSCAN, hierarchical
- Validation
 - Cross-validation, scoring, ROC
- Misc
 - PCA, MDS, item set mining
- External
 - R, Weka

55

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Visualization



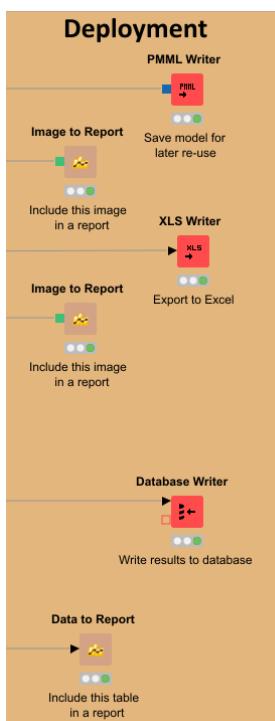
- Interactive
 - Scatter plot, histogram, pie charts, box plot
 - Highlighting (brushing)
- JFreeChart
- JavaScript
- Misc
 - Tag cloud, open street map, networks, molecules
- External
 - R

56

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Deployment

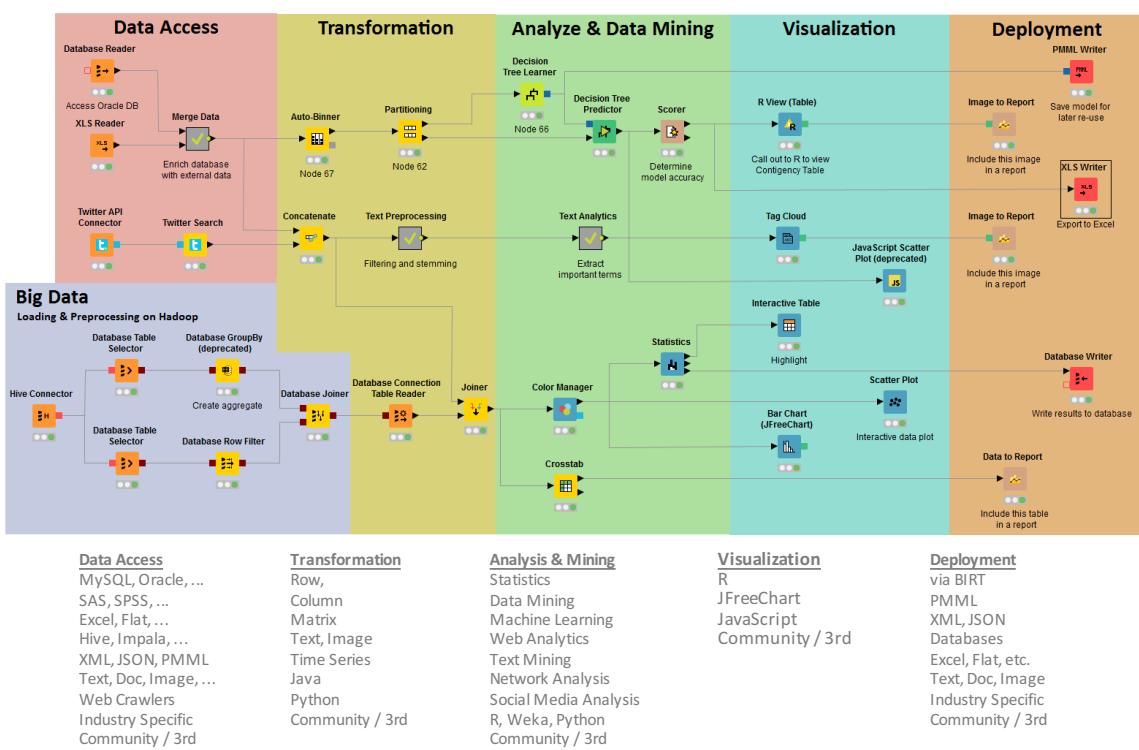


- Database
- Files
 - Excel, csv, txt
 - XML
 - PMML
 - to: local, KNIME Server, SSH-, FTP-Server
- BIRT Reporting

57

AdvancedAnalytics
.Academy

Over 1500 native and embedded nodes included:



58

This educational material was produced for the Machine Learning Workcamp at IWK Region Stuttgart December 2017. The copyright is with AdvancedAnalytics.Academy GmbH, Germany. Do not copy or distribute.

AdvancedAnalytics
.Academy

Overview

- Installing KNIME Analytics Platform
- The KNIME Workspace
- The KNIME File Extensions
- The KNIME Workbench
 - Workflow editor
 - Explorer
 - Node repository
 - Node description
 - Preferences
- Installing new features

59

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Install KNIME Analytics Platform

- Select the KNIME version for your computer:
 - Mac, Win, or Linux and 32 / 64bit
- Note different downloads (minimal or full)

- Download archive and extract the file, or download installer package and run it

Windows	
KNIME Analytics Platform for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	32 bit (301,82 MB) 64 bit (305 MB)
KNIME Analytics Platform + all free extensions for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	32 bit (1,74 GB) 64 bit (1,92 GB)
KNIME Analytics Platform for Windows (self-extracting archive) <i>The self-extracting archive only creates a folder holding the KNIME installation</i>	32 bit (304,18 MB) 64 bit (306,45 MB)
KNIME Analytics Platform for Windows (zip archive)	32 bit (246,61 MB) 64 bit (349,93 MB)
Linux	
KNIME Analytics Platform for Linux	32 bit (363,02 MB) 64 bit (360,1 MB)
KNIME Analytics Platform + all free extensions for Linux	32 bit (1,66 GB) 64 bit (2,07 GB)
Mac OSX	
KNIME Analytics Platform for Mac OSX (10.7 and above)	64 bit (329,41 MB)
KNIME Analytics Platform + all free extensions for Mac OSX (10.7 and above)	64 bit (1,99 GB)

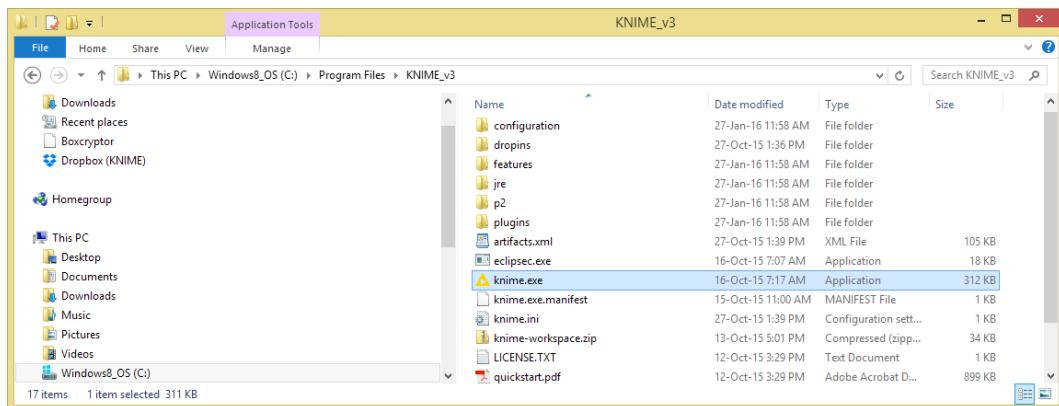
60

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Start KNIME Analytics Platform

- Go to the installation directory and launch **KNIME**, or use the shortcut created on your Desktop.



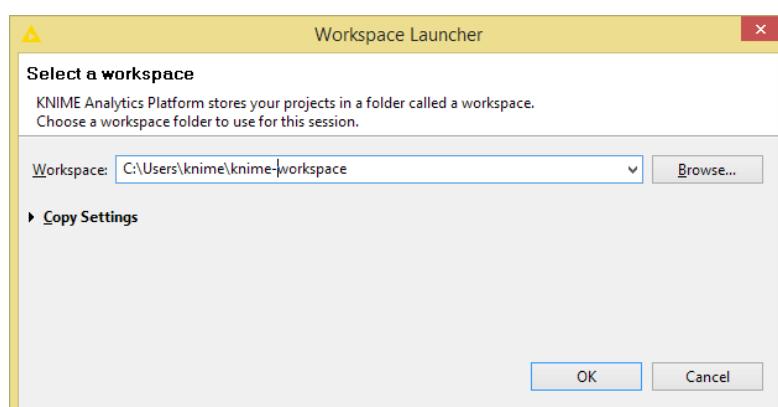
61

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



The KNIME Workspace

- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session.
- Workspaces are portable (just like KNIME)



62

This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute.



Welcome Page

The screenshot shows the KNIME Analytics Platform's welcome page. At the top left is the KNIME logo with the tagline "Open for Innovation". Below the logo is a banner with the text "Welcome to KNIME Analytics Platform!". The main content area includes:

- A section titled "New to KNIME? Looking for resources to get started?" with links to register for emails, explore the Quickstart Guide, check out 7 things to do after installing, find more hints and how-tos in the Learning Hub, and register for release and event emails.
- A note stating "This page will be displayed upon startup but you can customize the content using the checkboxes at the bottom."
- A section titled "Updates for the following components are available:" listing DYMATRUX Uplift Modeling Extensions and Palladian for KNIME, with a link to click here to install updates.
- A "Workflow Coach" section showing a list of recommended nodes: Community (8%), Column Filter (8%), Partitioning (8%), Joiner (8%), and k-Means (6%). A yellow arrow points from the "Partitioning" node in the list to its corresponding node in a workflow diagram.
- A "Where to go from here" sidebar with links to Create new workflow, Learning Hub, Browse example workflows, Get additional nodes, Go to my workflows, and Mount KNIME Cloud Server.
- A "Most recently used workflows" sidebar listing ModelSelection_WebPortal_Part1, ModelSelection_WebPortal_Part1, ModelSelection_BasicWorkflow, DataCleaning_WebPortal_v2.0, KNIME_project2, and Sexy ETL_v2.0.
- A "Tips & Tricks" section with a "Specialist Nodes" heading and a note about the variety of specialist nodes available.
- A checkbox section for "Show intro text at next start", "Show update notifications at next start", and "Show links and most recently used workflows at next start".
- A footer note: "This educational material was produced for the Machine Learning Workcamp at IHK Region Stuttgart December 2017. The copyright is with AdvancedAnalyticsAcademy GmbH, Germany. Do not copy or distribute."
- A footer logo for AdvancedAnalytics.Academy with the text "AdvancedAnalytics.Academy".

KNIME File Extensions

- Dedicated file extensions for Workflows and Workflow groups associated with KNIME Analytics Platform

- *.knwf for KNIME Workflow Files



- *.knar for KNIME Archive Files





AdvancedAnalytics
.Academy

www.advancedanalytics.academy

Thank you for your
attention!

Contact

Stefan Weingaertner
CEO

AdvancedAnalytics.Academy GmbH
E. sw@advancedanalytics.academy
T. +49 711 658 238 80
F. +49 711 658 238 88
M. +49 160 55 63 811