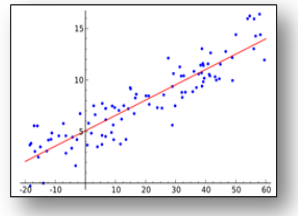
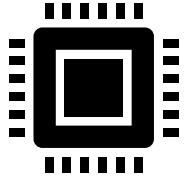
RAAK Top-up AIoTValley

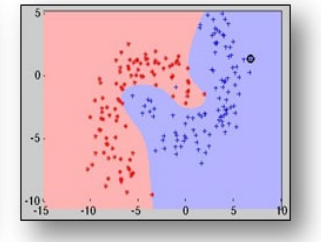
Artificial Intelligence Training session 2
Machine learning

Miha Lavrič

2021-06-04







SAXION UNIVERSITY OF APPLIED SCIENCES

TVALLEY

BOOST

Research projects

Focus op Vision (2019 – 2021)
- Enabling companies in ‘maakindustrie’ to use computer vision
- Disseminating knowledge on (AI for) vision

Data in Smart Industry (2017 – 2020)
- Enabling companies in ‘maakindustrie’ to use data for process optimization
- Disseminating knowledge on IoT and AI for data acquisition and analysis

AIoTValley (2020 – 2021)
- ‘Top-up’ of DSI to create open access educational material
- Including AI and IoT in TValley (robotics & mechatronics fieldlab)

Introduction square

Motivation

RAAK projects + BOOST

Involved partners

Learning goals

Lights, camera and “coffee”…

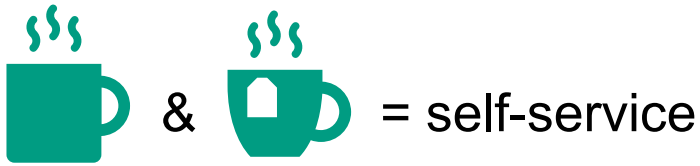
Webinar, but interactive

Ask questions in chat

Assisted by Jeroen Linssen

Max. 3 hours with 2 breaks

Session is being recorded

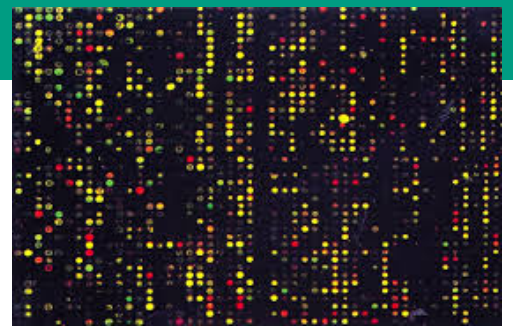


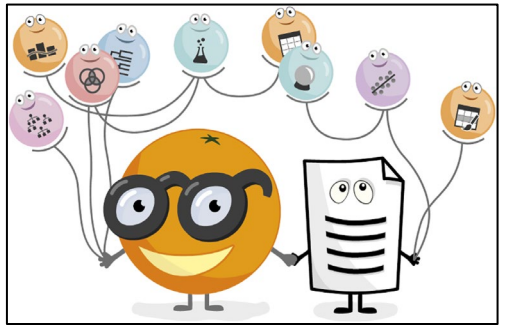
Your learning goals and teacher for today…

- Get to know machine learning
  - Principles
  - Algorithms
  - Evaluation
- Demonstration of ML in Orange
- Hands-on experience in ML
  - Visual programming – Orange
  - Coding – Python in Colab

Miha Lavrič

Researcher / Lecturer Applied Data Science
Ambient Intelligence
Saxion





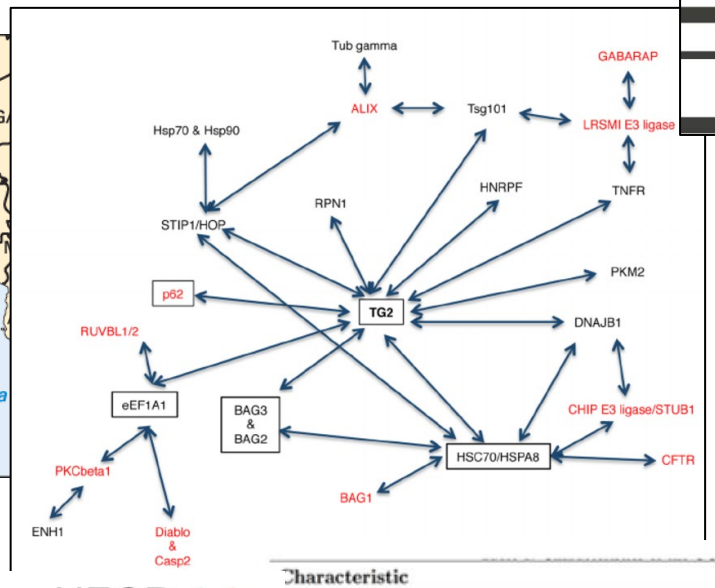
SLOVENIA

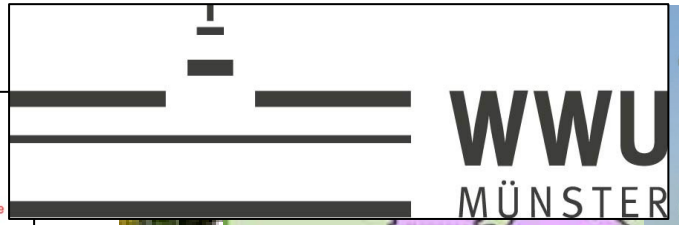
DŠMBS

Univerza v Ljubljani

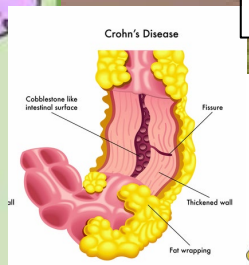
SAXION UNIVERSITY OF APPLIED SCIENCES

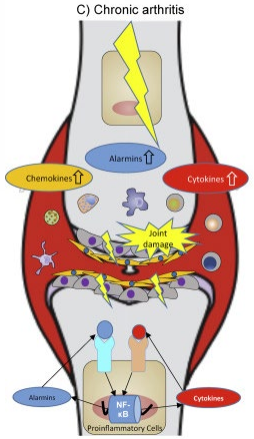
GLUTEN FREE

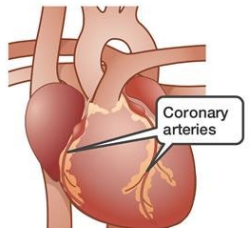


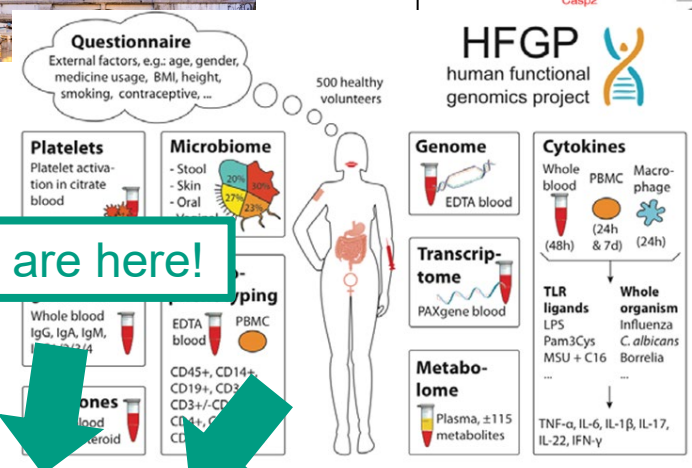


UKM
Universitätsklinikum Münster





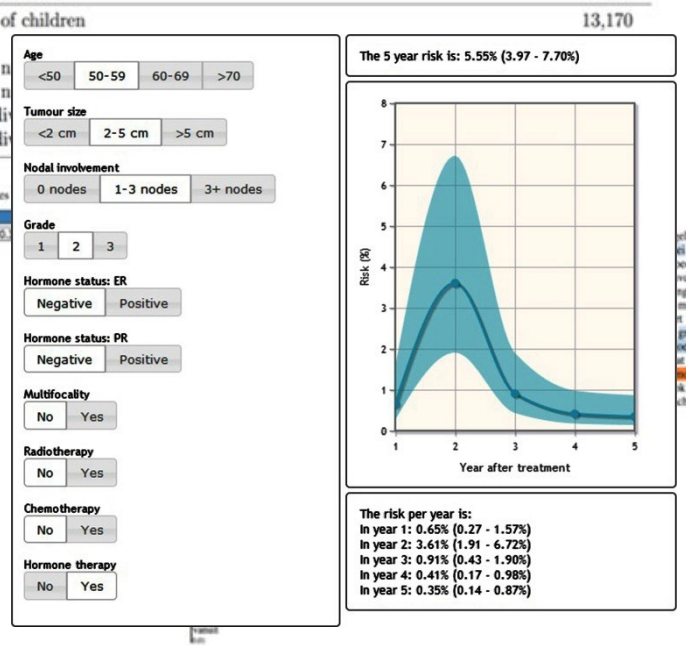


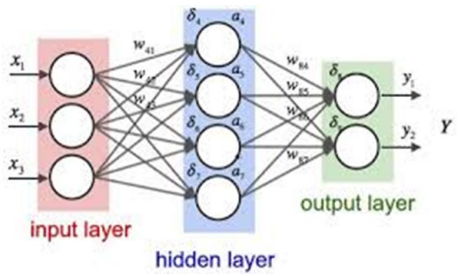


You are here!

HFGP
human functional genomics project

Characteristic	Value
Number of children	13,170
Consults	
Average n	
Average n	
Lexical di	
Lexical di	

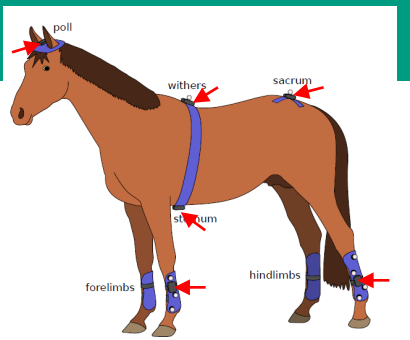




SAXION
UNIVERSITY OF APPLIED SCIENCES

MRDM
Medical Research Data Management

My past and present work at Saxion…



Hanzehogeschool Groningen
University of Applied Sciences

umcg

SAXION

Roessingh
Centrum voor Revalidatie

de Fysiotherapeut
Nederlandse Vereniging voor Fysiotherapie in de Sportgezondheidszorg

EUROSPED
TopVolleybal in Twente

360 sports intelligence

FysioCentrum Nijverdal

TopvormTwente.

CE-Mate

Big data technologie voor detectie overbelasting sporters

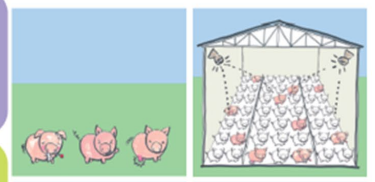
Waarom nu?
EFRO x3D
Digitaal (x) monitoren van:
- Diergedraag
- Diergezondheid
- Dierenwelzijn

Het x3D-project richt zich op dieren waar samenwerking tussen mens en dier intens is (paarden) en op individuele in groepen gehouden dieren die voor ons voedsel zorgen (zoals koeien, kippen, varkens, geiten en schapen).

Kerndoelen:
- Inspireren
- Verbinden
- Simuleren
- Kansen creëren en realiseren

Europese Unie

OPoost

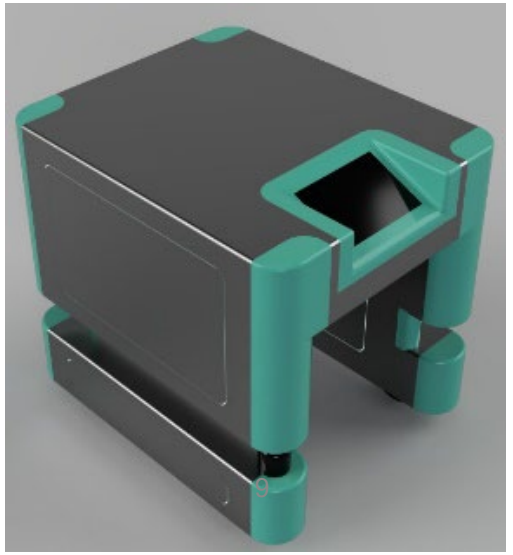


WAGENINGEN UNIVERSITY & RESEARCH

UNIVERSITY OF TWENTE.

van hall larenstein university of applied sciences

SAXION



Miha's Magical Machine Learning "Masterclass" for SMEs

Overview

1. **Machine learning**
   a) Principles
   b) Evaluation
   c) Algorithms (supervised learning / classification / regression)

2. **ML with Orange**

3. **Hands-on ML with Python in Colab**

4. **Take home message, recap & outlook**

And now for some fun!

$$Y_i = \beta_0 + \beta_1 X_i + \acute{\varepsilon}_i$$

$$P = \frac{e^{a+bX}}{1+e^{a+bX}}$$

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times w_{ij}\right)^2$$

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

$$\text{maximize } f(c_1 \ldots c_n) = \sum_{i=1}^{n} c_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i (\varphi(\vec{x}_i)\cdot\varphi(\vec{x}_j)) y_j c_j$$
$$= \sum_{i=1}^{n} c_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i k(\vec{x}_i, \vec{x}_j) y_j c_j$$

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

$$K_k^{cc}(\mathbf{x},\mathbf{z}) = \sum_{k_1,\ldots,k_d,\sum_{j=1}^{d} k_j = k} \frac{k!}{k_1!\cdots k_d!}\left(\frac{1}{d}\right)^k \prod_{j=1}^{d} \mathbf{1}_{\lceil 2^{k_j} x_j\rceil = \lceil 2^{k_j} z_j\rceil},$$
$$\text{for all } \mathbf{x},\mathbf{z} \in [0,1]^d.$$

$$L(f_m) \approx \sum_{i=1}^{n}[g_m(x_i)f_m(x_i) + \frac{1}{2}h_m(x_i)f_m(x_i)^2] + const.$$
$$\propto \sum_{j=1}^{T_m}\sum_{i\in R_{jm}}[g_m(x_i)w_{jm} + \frac{1}{2}h_m(x_i)w_{jm}^2].$$

$$z = x_1 * w_1 + x_2 * w_2 + \ldots\ldots + x_n * w_n + b * 1$$
$$\hat{y} = a_{out} = sigmoid(z)$$
$$sigmoid(z) = \frac{1}{1+e^{-z}}$$

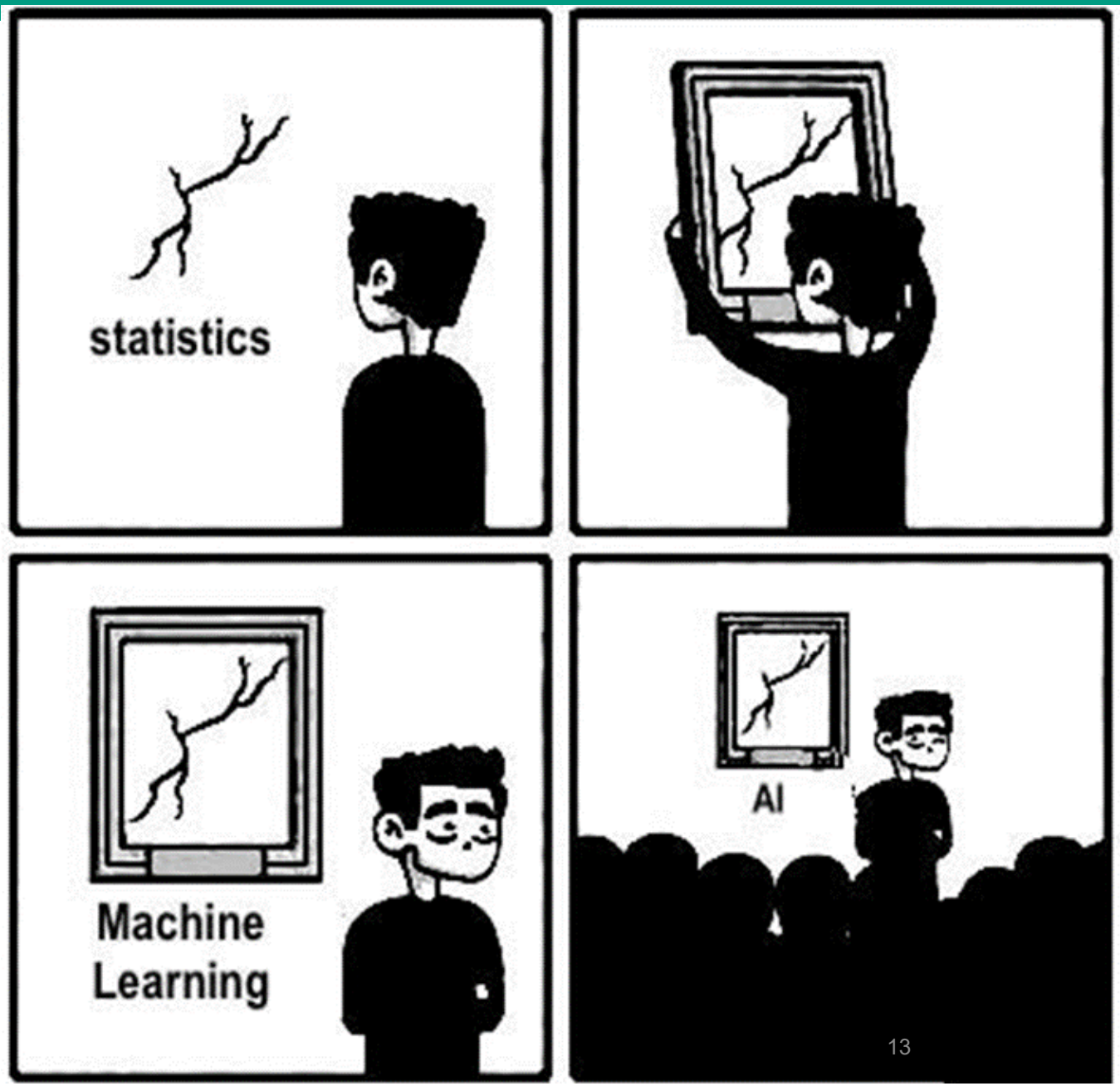
$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

$$\sum_{i=1}^{k}|x_i - y_i|$$

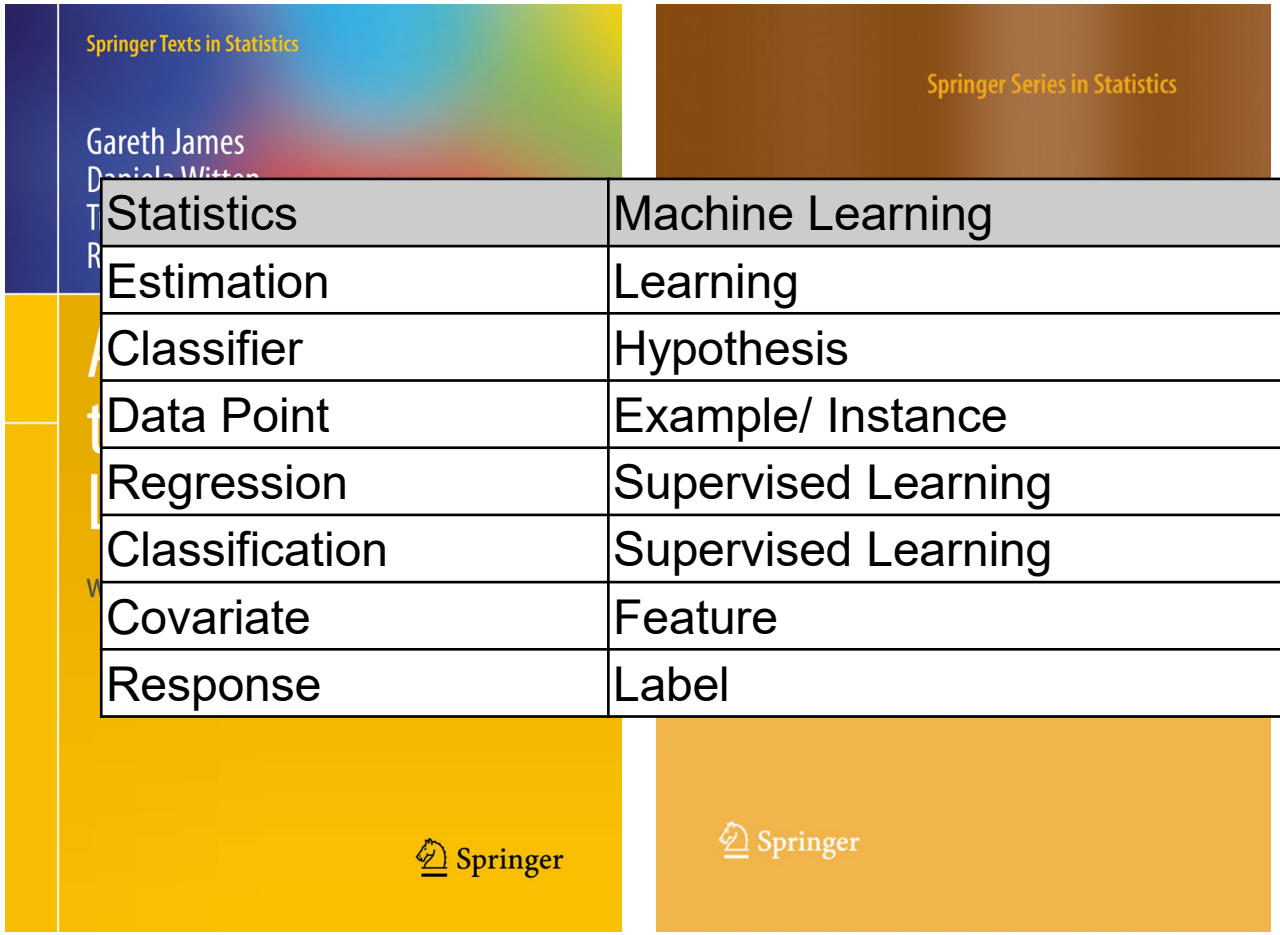
$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

"When you're fundraising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression."
—everyone on Twitter ever

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$



Robert Tibshirani, a statistician and machine learning expert at Stanford, called machine learning “glorified statistics."



Statistics	Machine Learning
Estimation	Learning
Classifier	Hypothesis
Data Point	Example/ Instance
Regression	Supervised Learning
Classification	Supervised Learning
Covariate	Feature
Response	Label

Glossary

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = $1,000,000	large grant= $50,000
nice place to have a meeting: Snowbird, Utah, French Alps	nice place to have a meeting: Las Vegas in August

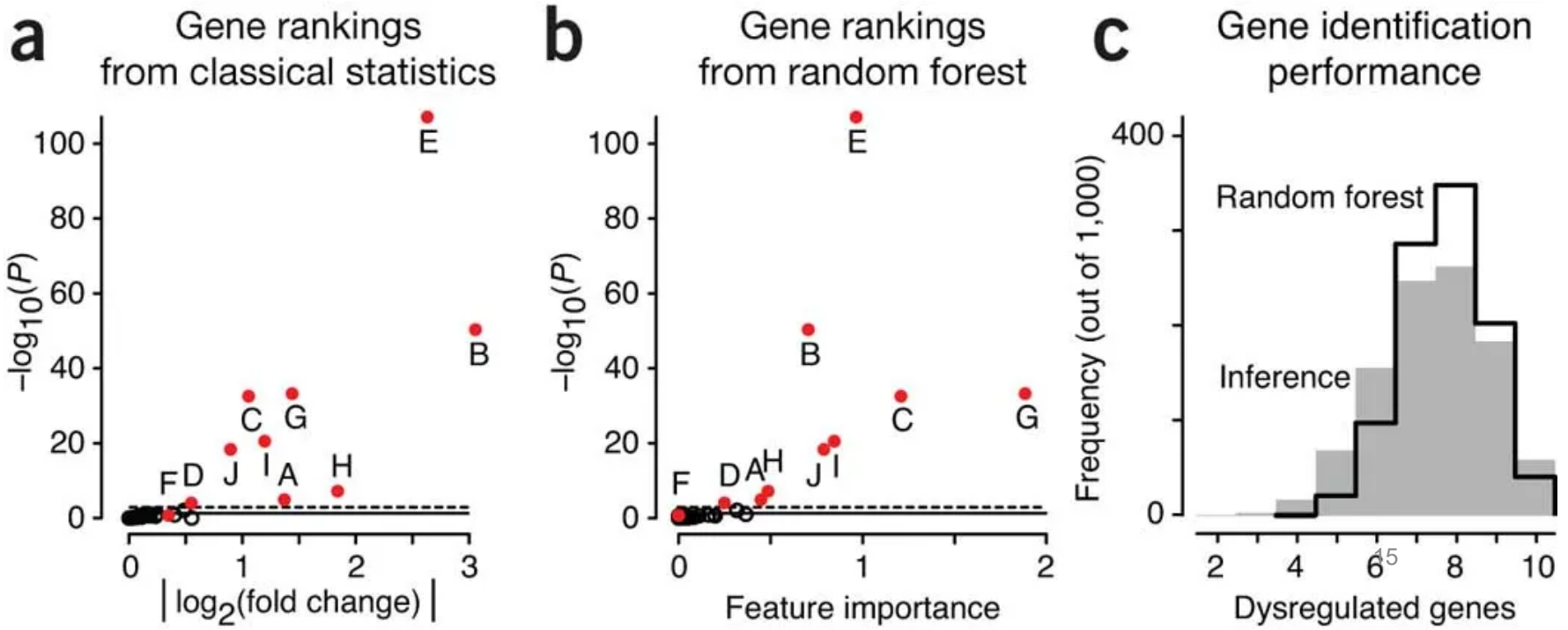
The boundary between statistical inference and ML is subject to debate[1]—some methods fall squarely into one or the other domain, but many are used in both. …

Statistics requires us to choose a model that incorporates our knowledge of the system, and

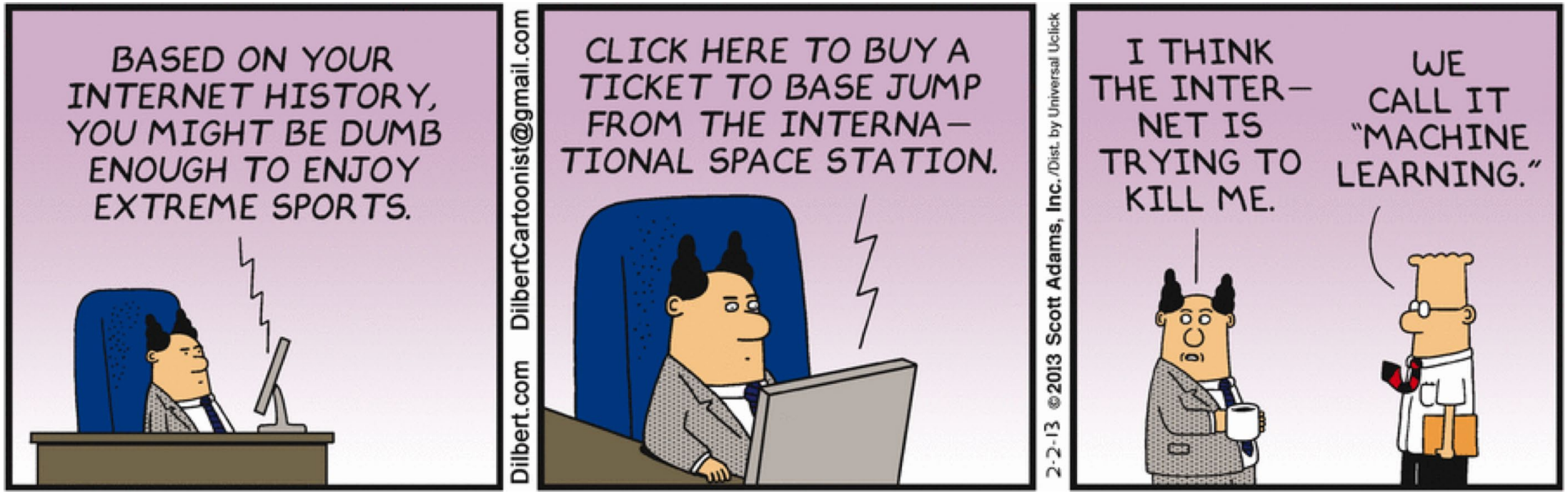
ML requires us to choose a predictive algorithm by relying on its empirical capabilities. …

Inference and ML are complementary in pointing us to biologically meaningful conclusions.

*Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. Nat Methods **15,** 233–234 (2018).*

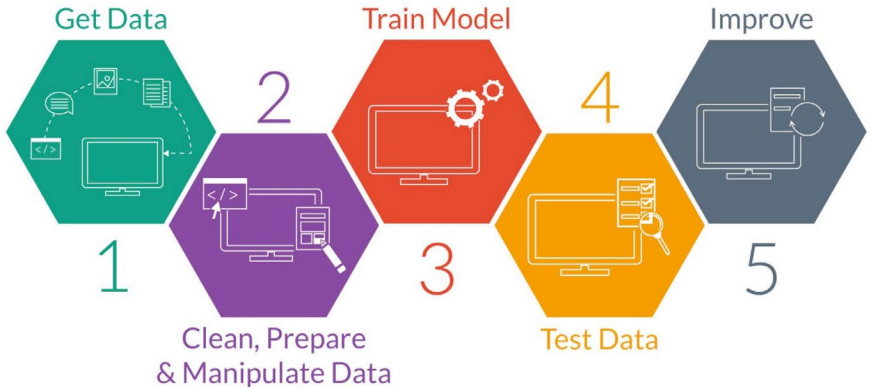


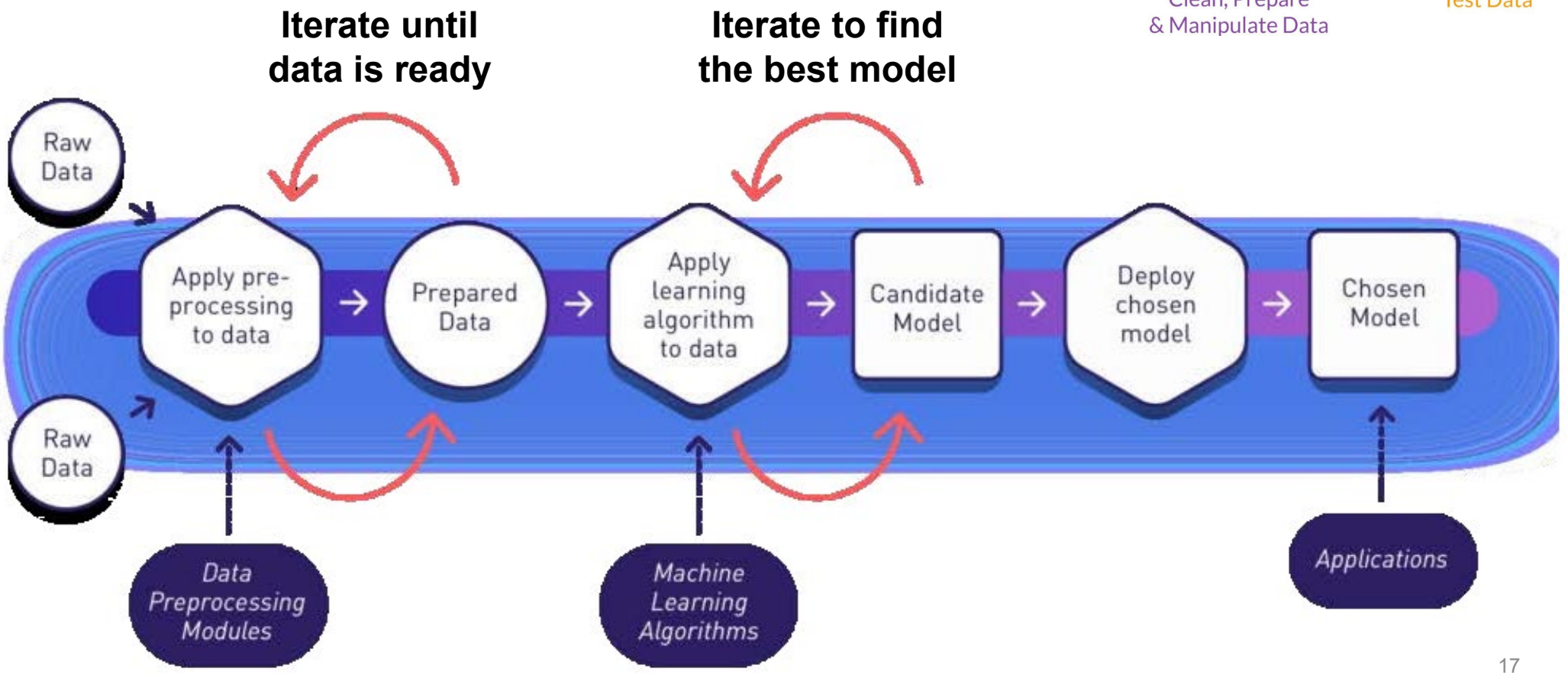
What is machine learning?



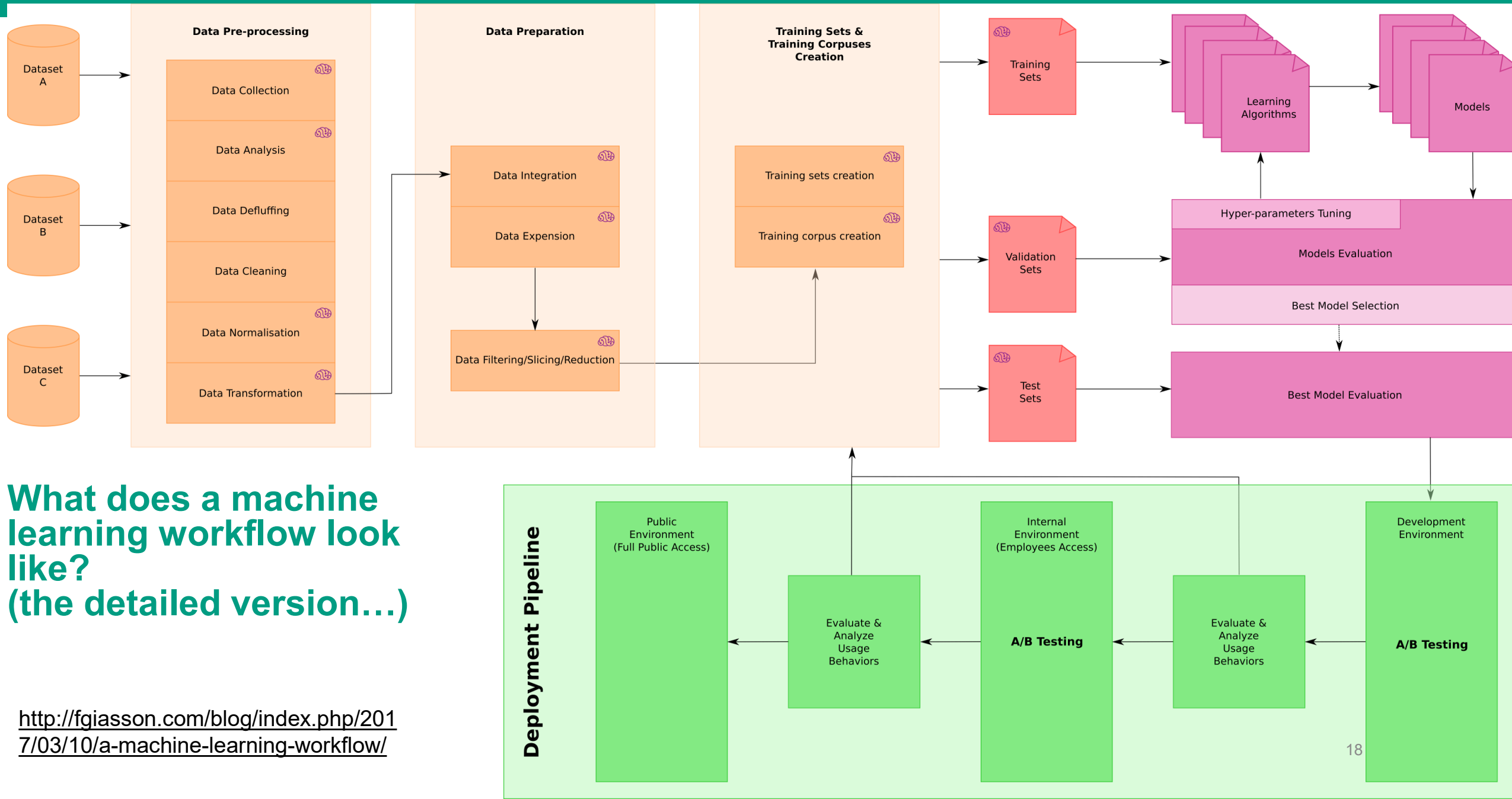
- A field of **computer science / artificial intelligence** that uses **statistical techniques** to give computer systems the ability to "learn" (e.g. progressively improve performance on a specific task) from data, without being explicitly programmed…

What does a machine learning workflow look like? (the simple version)



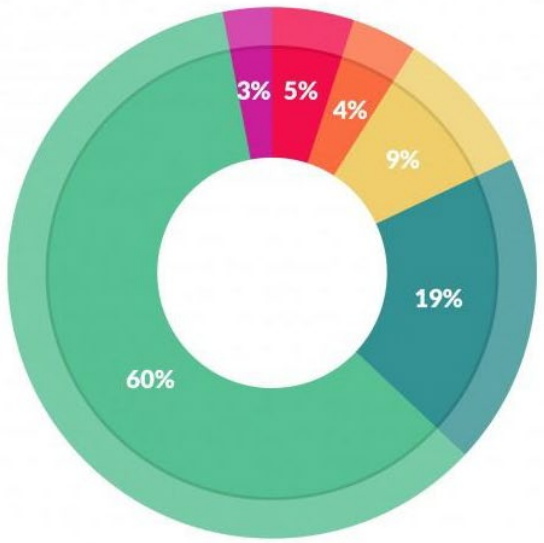


What does a machine learning workflow look like?
(the detailed version…)



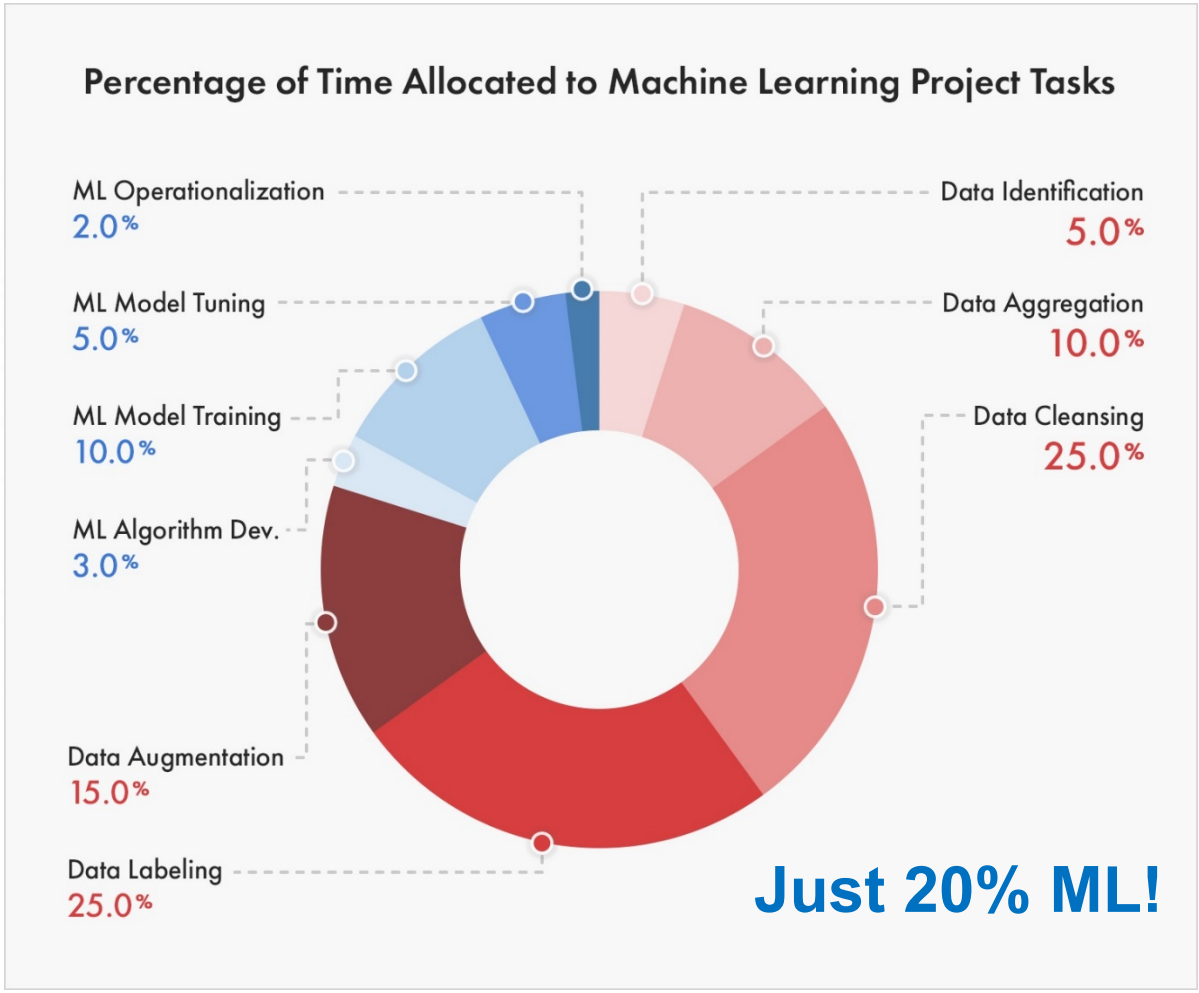
http://fgiasson.com/blog/index.php/2017/03/10/a-machine-learning-workflow/

How do machine learning scientists spend their time?



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

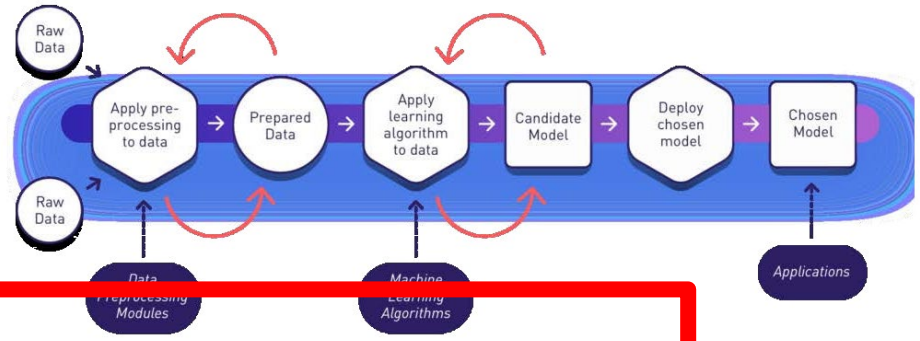


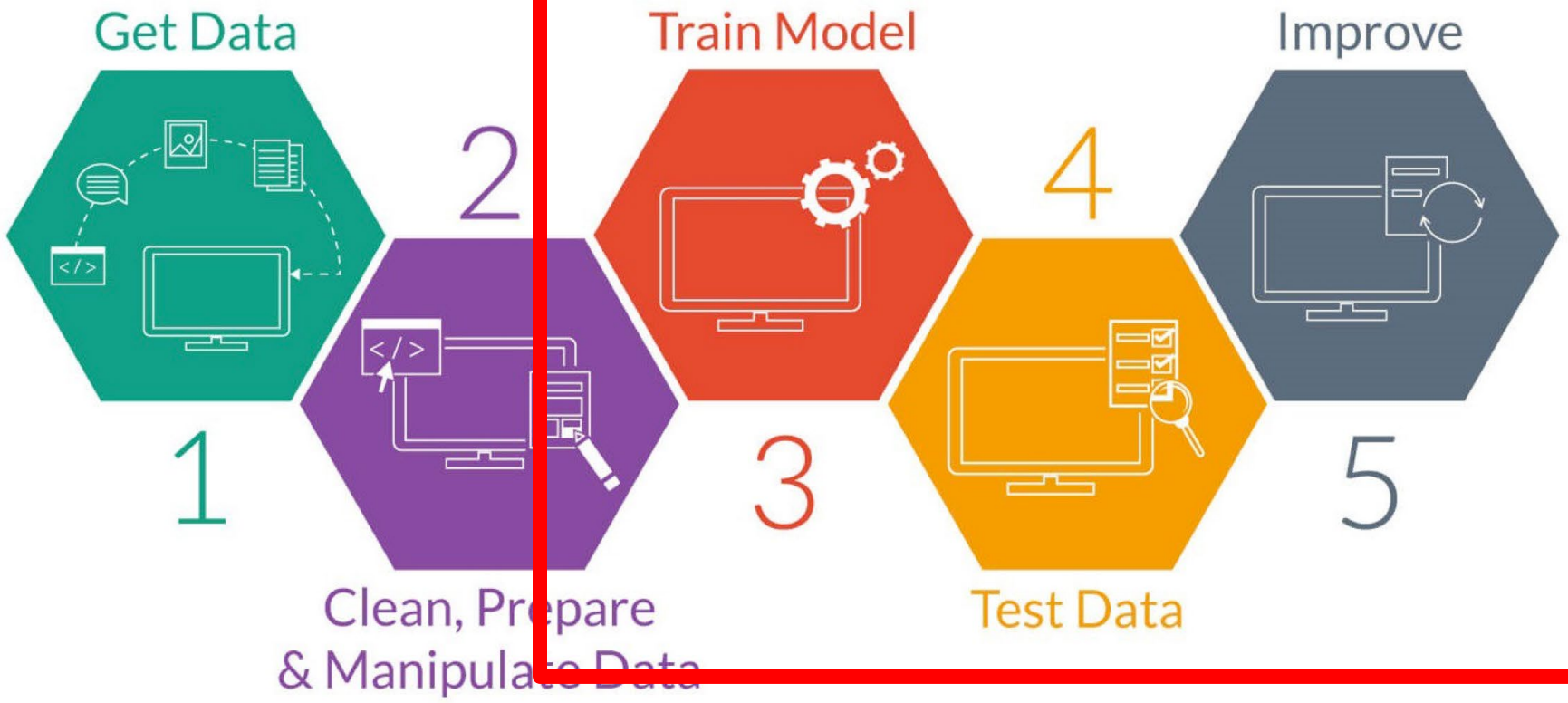
https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

https://www.cloudfactory.com/data-labeling-guide

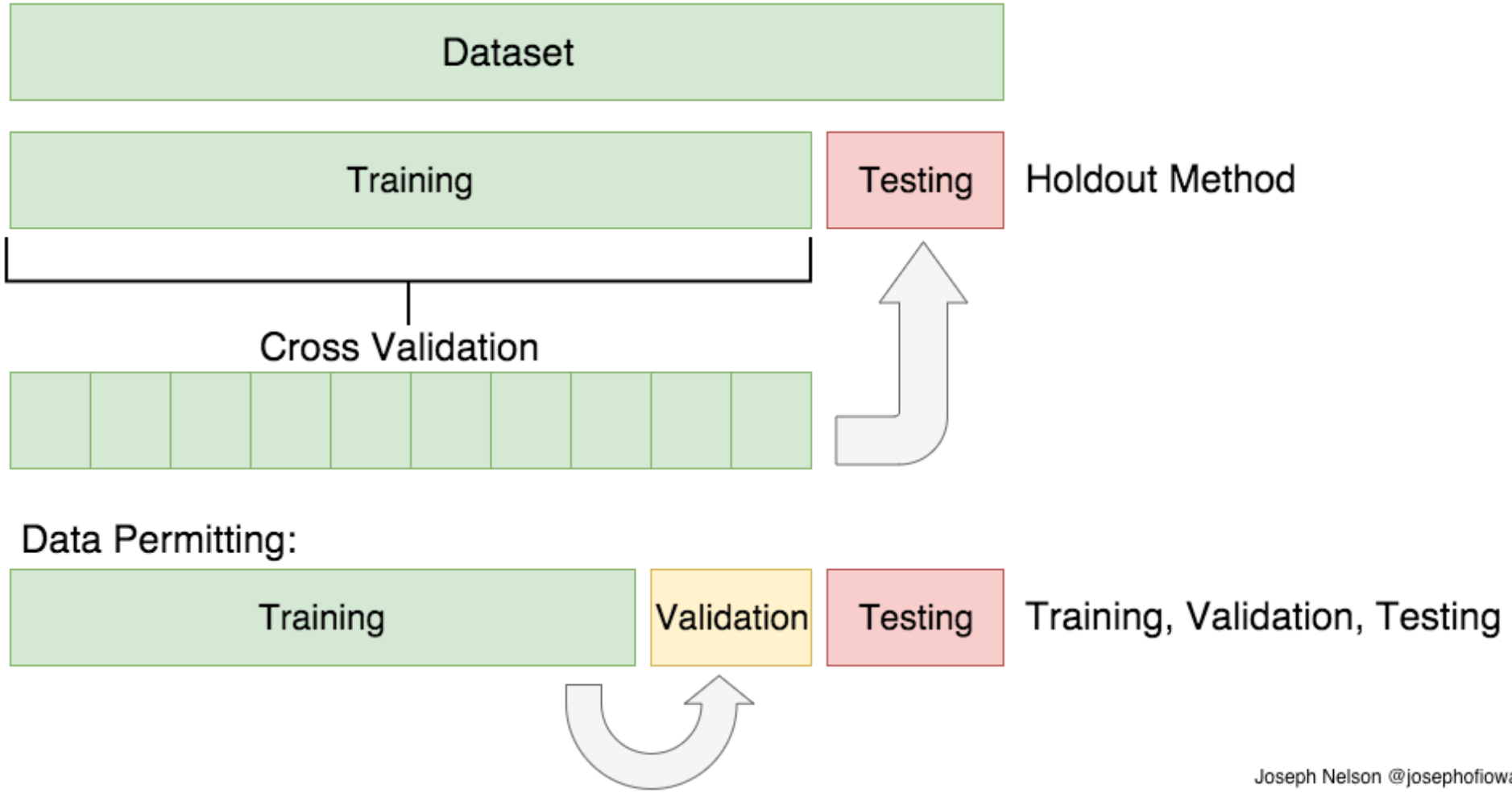
SAXION UNIVERSITY OF APPLIED SCIENCES

Today's focus in the practical sessions?



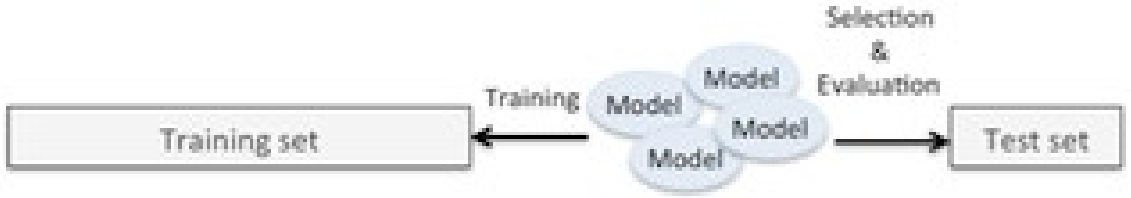


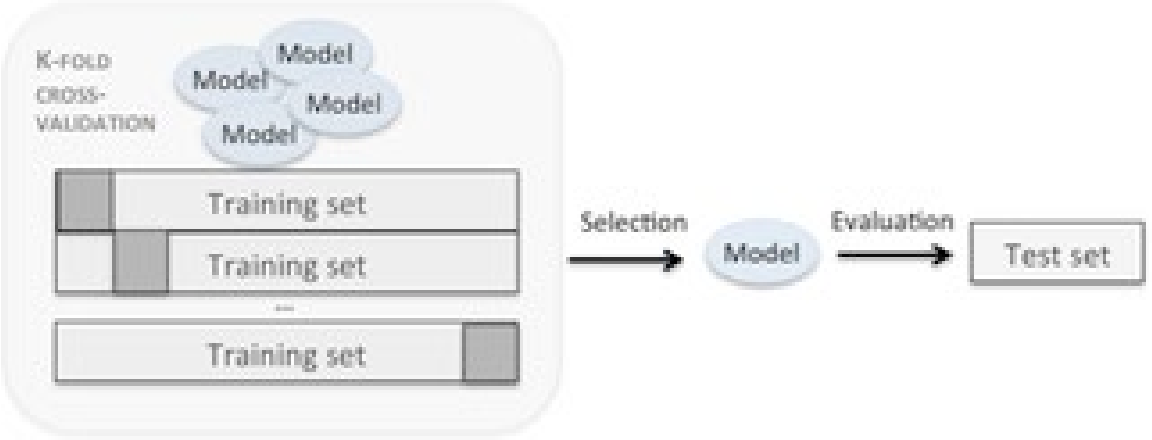
So we have some nice squeaky clean data…Now what? We split it!



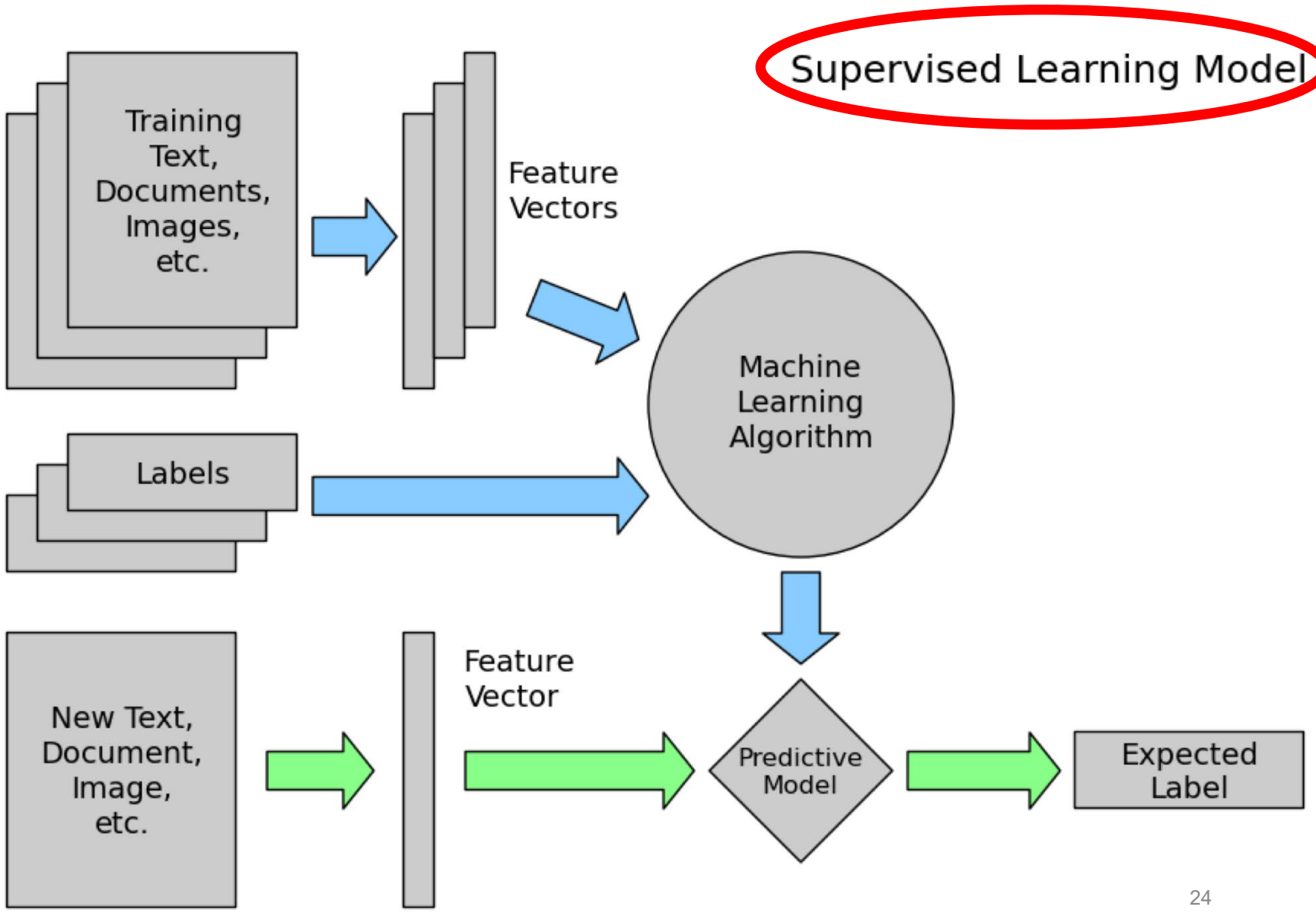
Joseph Nelson @josephofiowa

To look for the best model / algorithm…

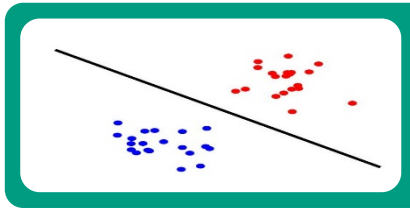




…and finally begin predicting!

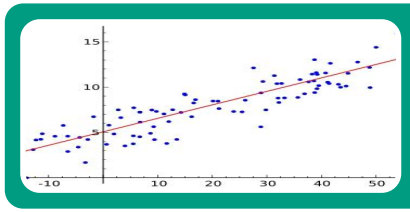


Modeling Techniques/ Methods



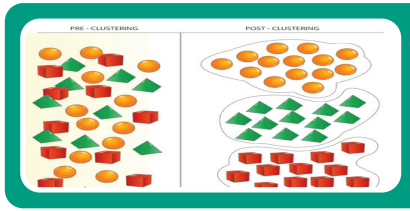
Classification: A model or classifier is constructed to predict class (categorical) labels

- Algorithms: Decision tree (DT), Bayesian, rule-based, support vector machines, artificial neural network (ANN), linear discriminant



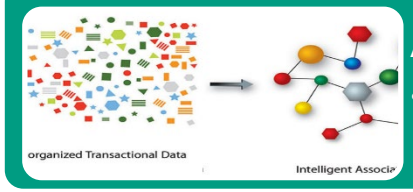
Prediction (Regression): A model performing prediction function to forecast future values of continuous type data

- Algorithms: Regression, ANN, support vector regression, DT, and Fuzzy set



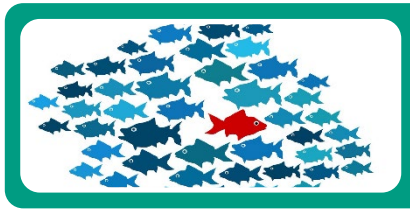
Clustering: Grouping a set of data objects into multiple clusters so that objects within a cluster have high similarity

- Algorithms: Centroid-based clustering, connectivity-based clustering, density-based clustering, and distribution-based clustering



Association: To discover interesting associations and correlations

- Algorithms: Apriori, AprioriAII, sampling, partitioning pattern growth, correlation rules, stream patterns



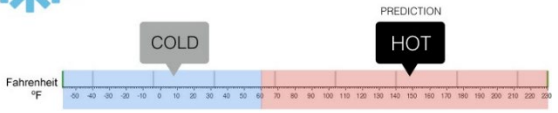
Anomaly/Outlier Detection: the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data

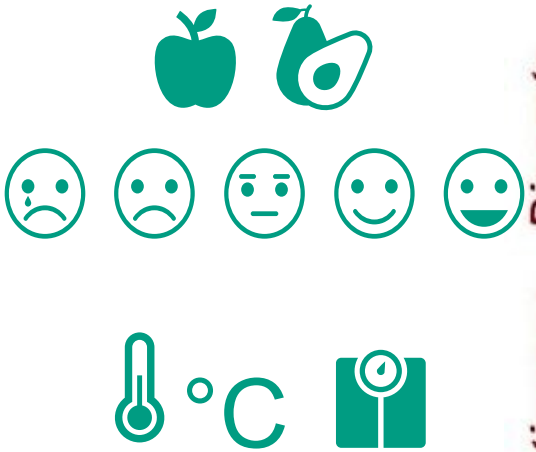
- Algorithms: k-nearest neighbour, local outlier factor, isolation forests, One-class support vector machines, Replicator neural networks

The main divides in ML approaches…

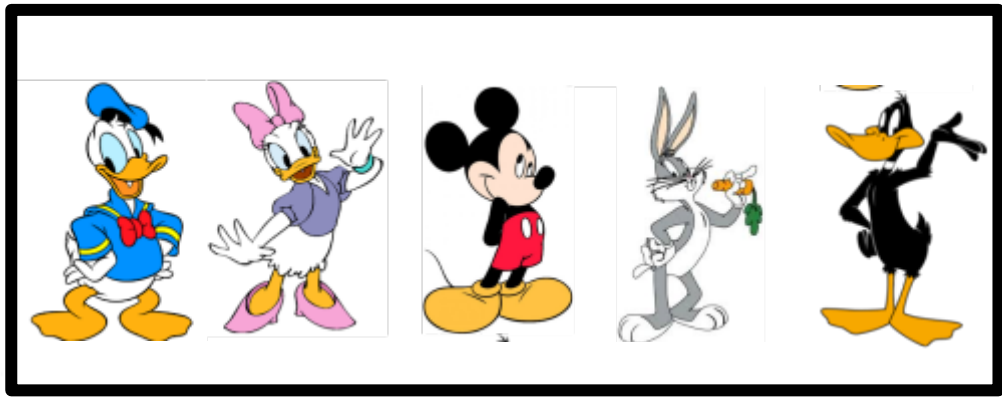
Classification

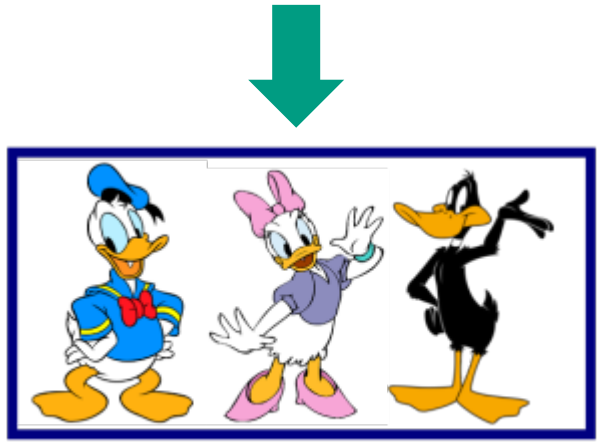
Will it be Cold or Hot tomorrow?

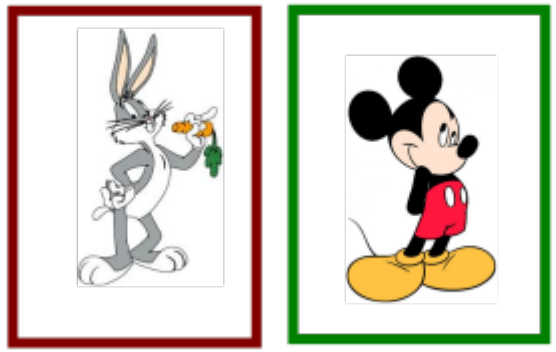




	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

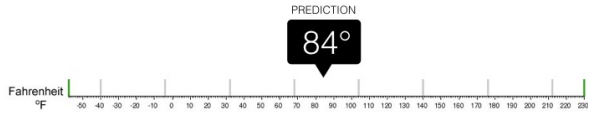




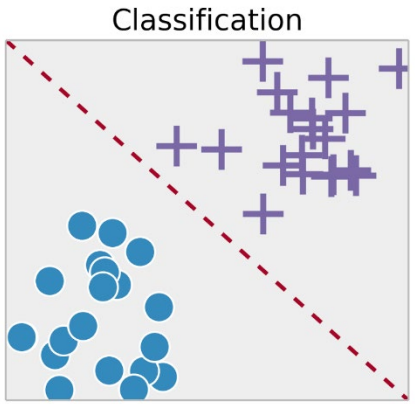


Regression

What is the temperature going to be tomorrow?

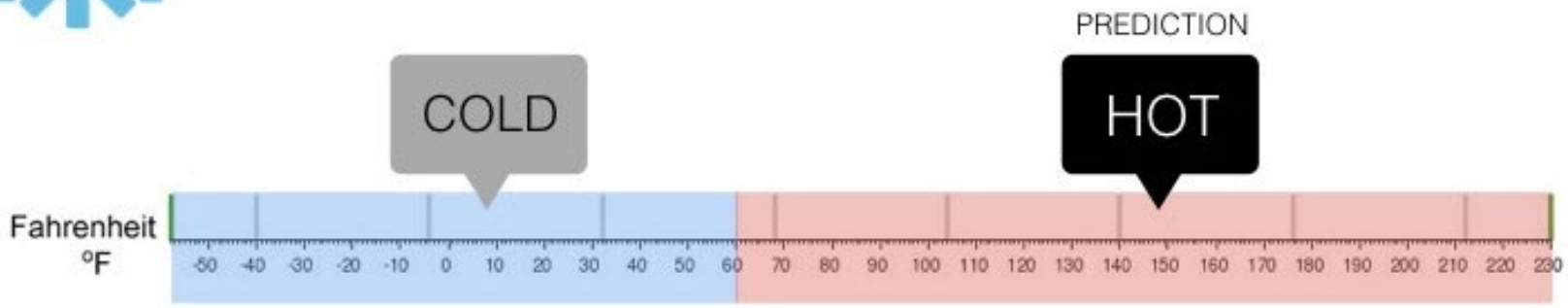


https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d



Classification

Will it be Cold or Hot tomorrow?

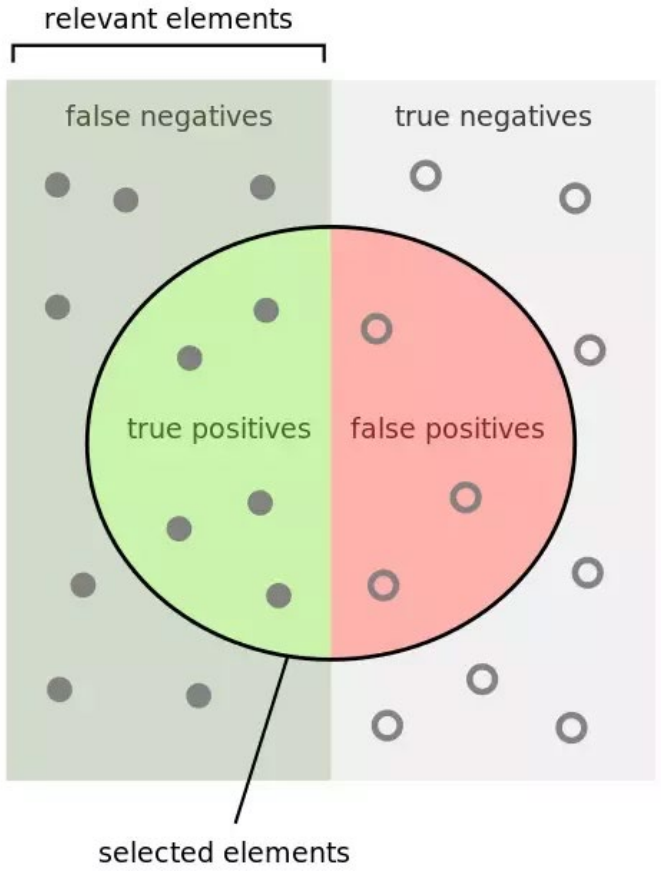


Nominal

Ordinal

A basic confusion matrix.

		Predicted class	
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

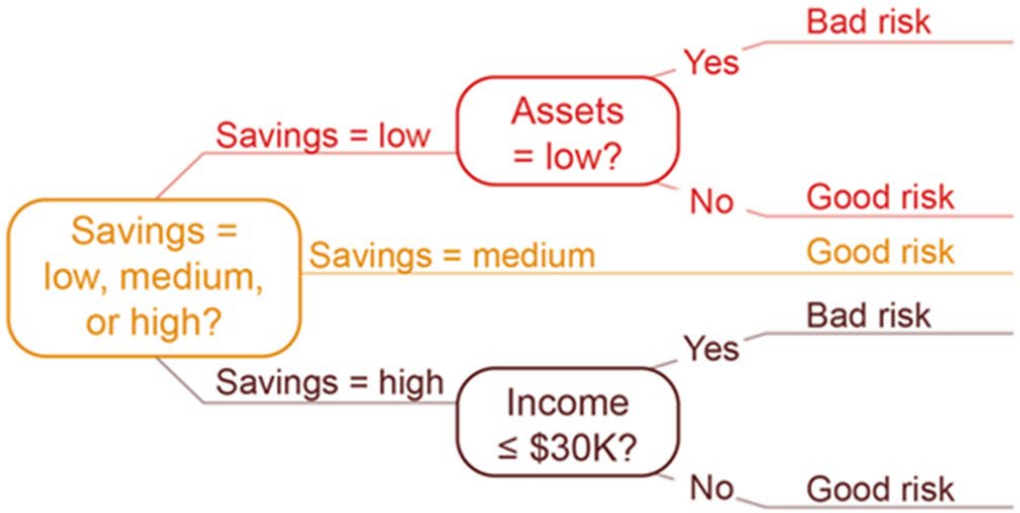
$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

AKA Sensitivity

F1-score

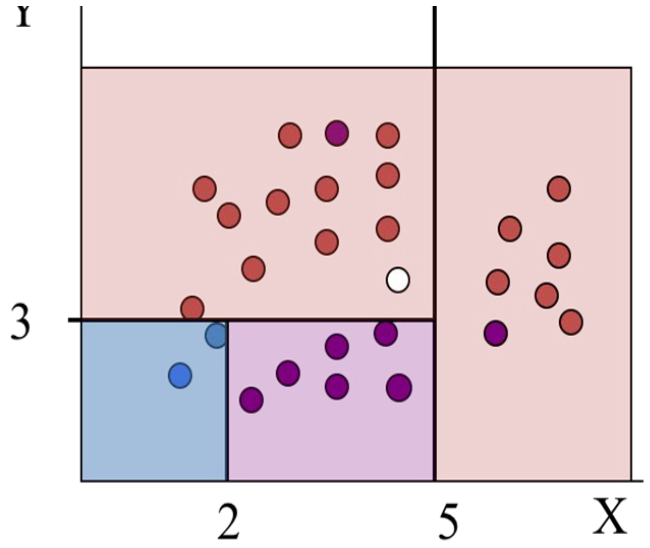
Decision trees

Decision tree analysis typically uses a hierarchy of variables or decision nodes that, when answered step by step, can classify a given customer as creditworthy or not, for example.



Advantages	Use cases
Decision trees are useful when evaluating lists of distinct features, qualities, or characteristics of people, places, or things.	Rule-based credit risk assessment, horse race performance prediction

Source: Daniel T. Larose and Chantal D. Larose, *Data Mining and Predictive Analytics*, 2nd Edition, John Wiley & Sons, 2015

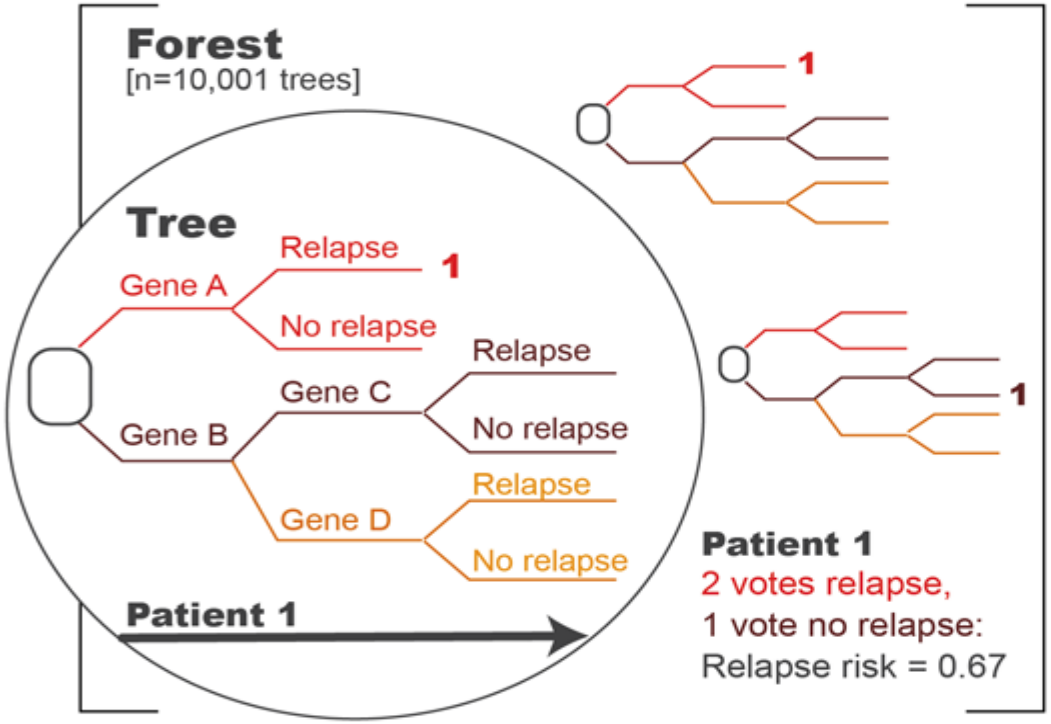


if X > 5 then orange
else if Y > 3 then orange
else if X > 2 then purple
else blue

SAXION UNIVERSITY OF APPLIED SCIENCES

Random forest

Random forest algorithms improve the accuracy of decision trees by using multiple trees with randomly selected subsets of data. This example reviews the expression levels of various genes associated with breast cancer relapse and computes a relapse risk.



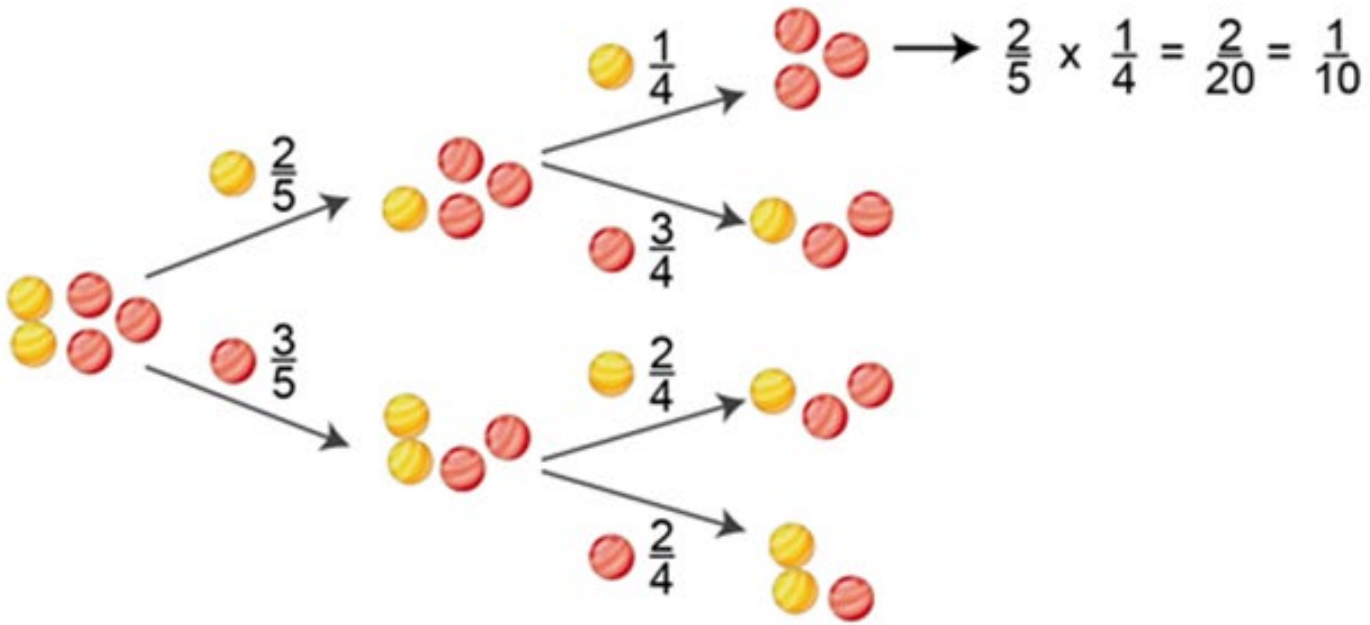
Advantages	Use cases
Random forest methods prove useful with large data sets and items that have numerous and sometimes irrelevant features.	Customer churn analysis, risk assessment

Source: Nicolas Spies, Washington University, 2015

An ensemble learning method that corrects for decision trees' habit of overfitting to their training set.

Naive Bayes classification

Naive Bayes classifiers compute probabilities, given tree branches of possible conditions. Each individual feature is "naive" or conditionally independent of, and therefore does not influence, the others. For example, what's the probability you would draw two yellow marbles in a row, given a jar of five yellow and red marbles total? The probability, following the topmost branch of two yellow in a row, is one in ten. Naive Bayes classifiers compute the combined, conditional probabilities of multiple attributes.

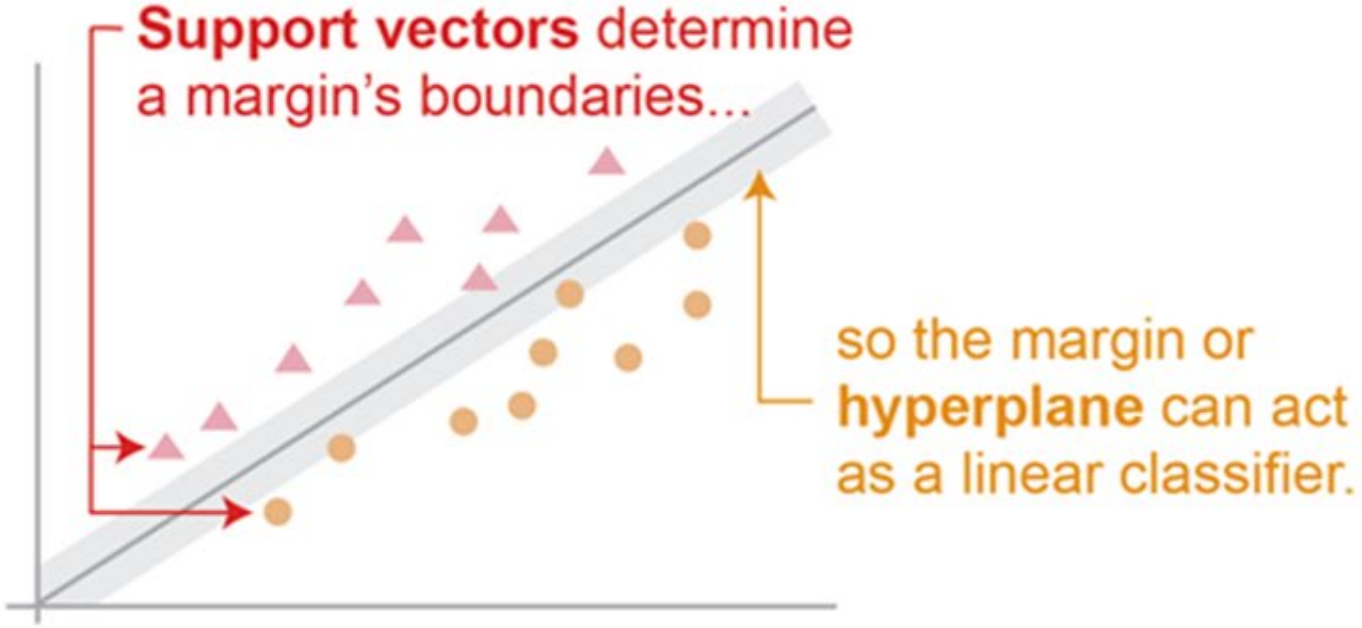


Advantages	Use cases
Naive Bayes methods allow the quick classification of relevant items in small data sets that have distinct features.	Sentiment analysis, consumer segmentation

Source: Rod Pierce, et al., *MathIsFun*, 2014

Support vector machines

Support vector machines classify groups of data with the help of hyperplanes.

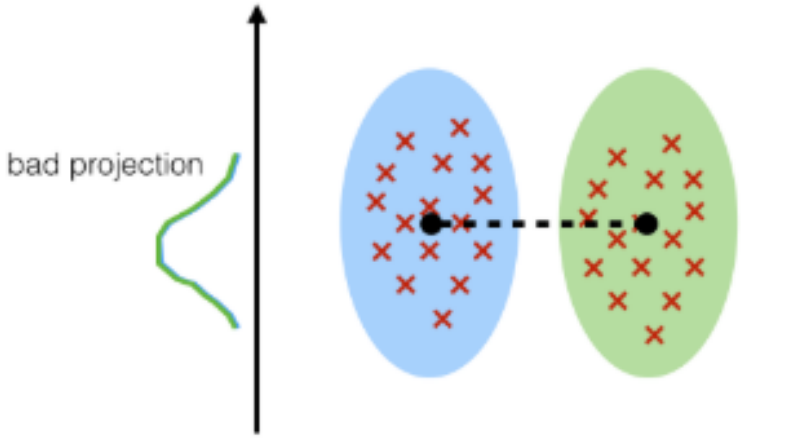


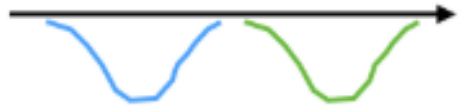
Advantages	Use cases
Support vector machines are good for the binary classification of X versus other variables and are useful whether or not the relationship between variables is linear.	News categorization, handwriting recognition

Source: Matthew Kelly, *Computer Science: Source*, 2010

Linear Discriminant Analysis

LDA:
maximizing the component axes for class-separation

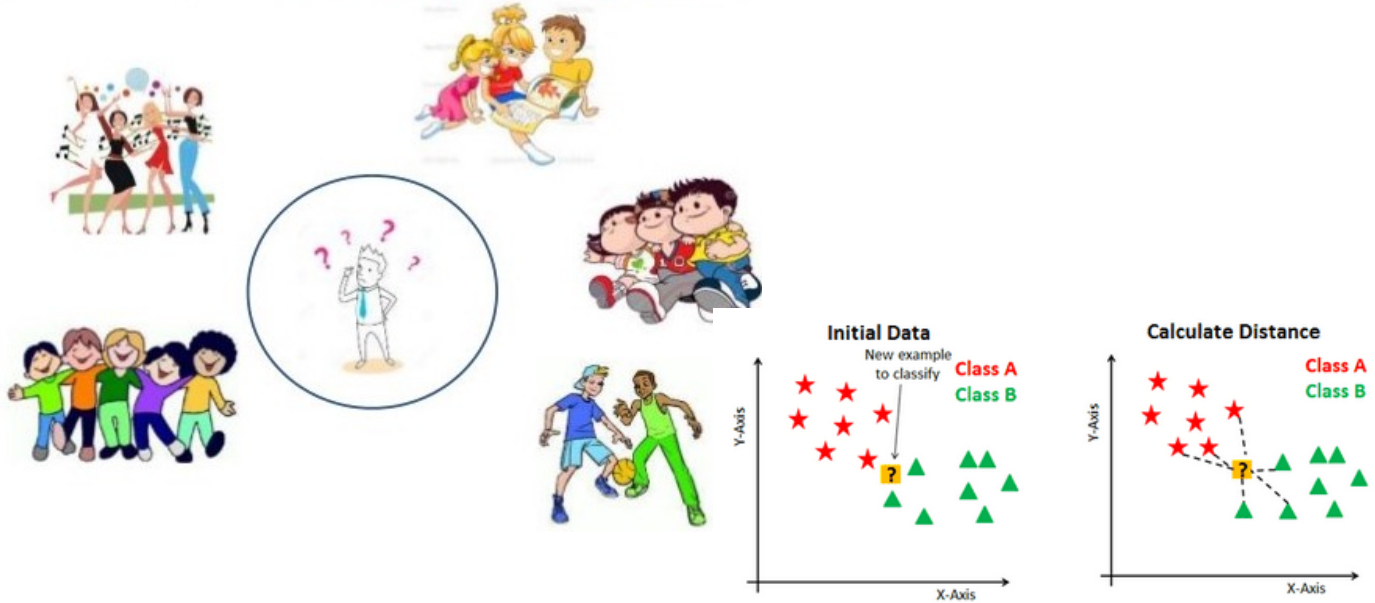


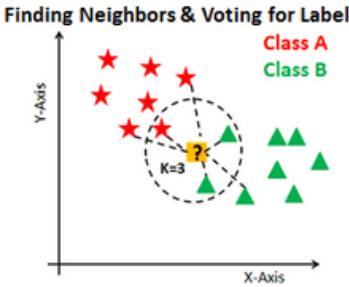


good projection: separates classes well

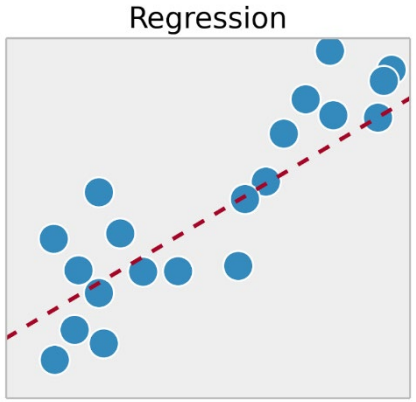
k-nearest neighbors algorithm (kNN)

Tell me about your friends(*who your neighbors are*) and *I will tell you who you are*.



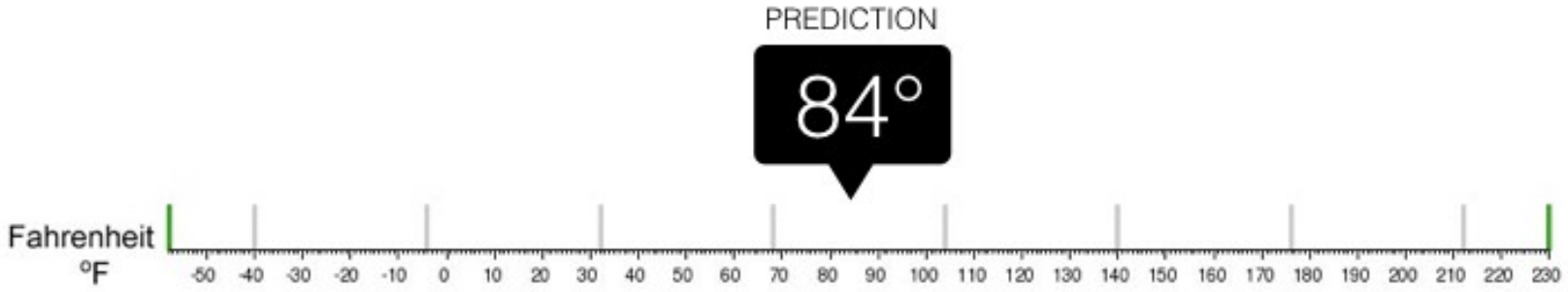


1. Divide the data into training and test data.
2. Select a value K.
3. Determine which distance function is to be used.
4. Choose a sample from the test data that needs to be classified and compute the distance to its n training samples.
5. Sort the distances obtained and take the k-nearest data samples.
6. Assign the test class to the class based on the majority vote of its k neighbors.



Regression

What is the temperature going to be tomorrow?



Regress

1. To return to a previous, usually worse or less developed state.
2. To have a tendency to approach or go back to a statistical mean.
3. To move backward or away from a reference point.

Origin:

- **Latin** *regressus,* a retreat,
- from *regredī,* to go back,
- from re- + *gradī,* to go,
- from *ghredh-* in **Indo-European roots,** meaning to walk, go.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

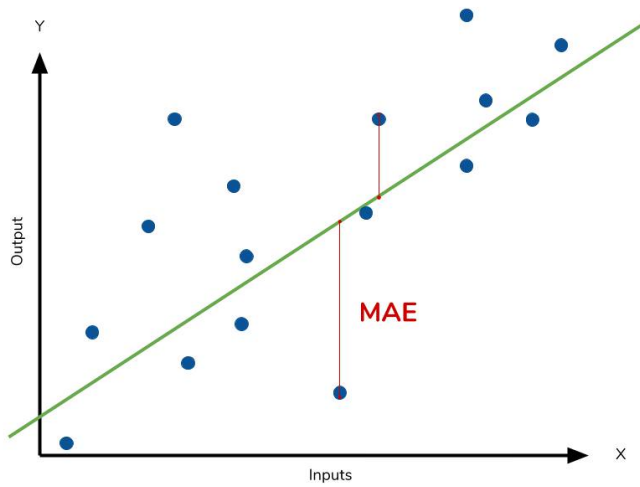
$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

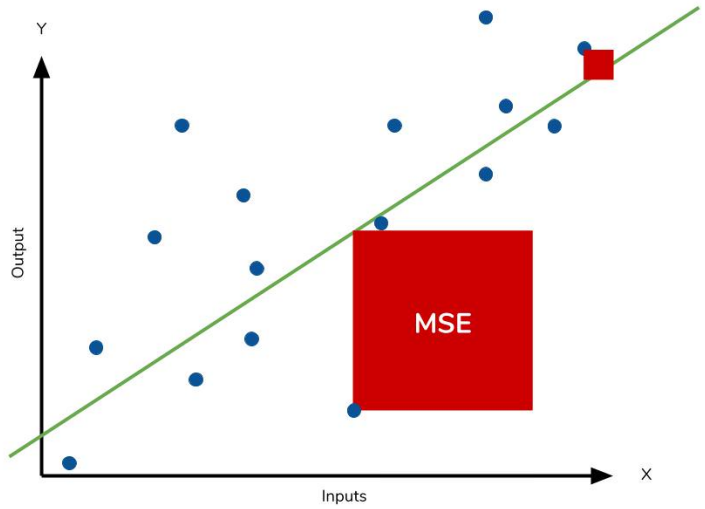
$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

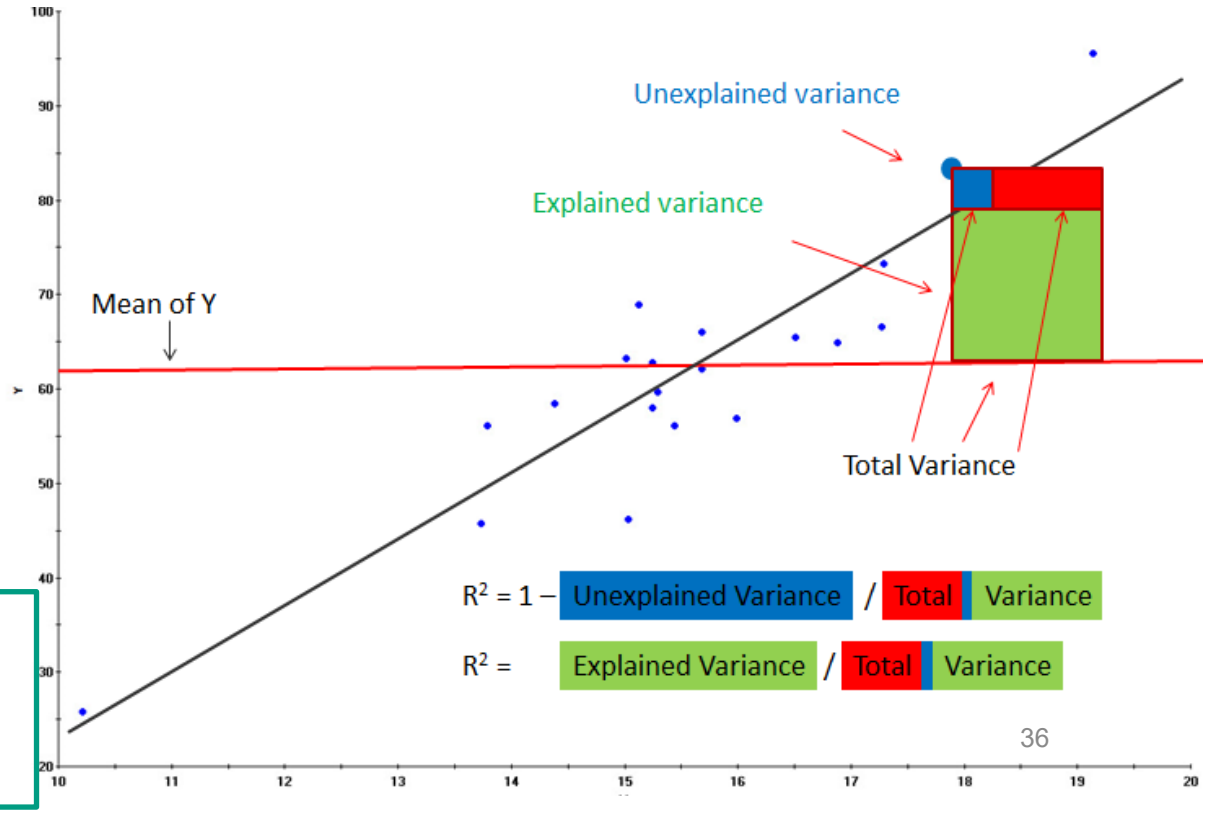
Where,

$\hat{y}$ – *predicted value of y*

$\bar{y}$ – *mean value of y*



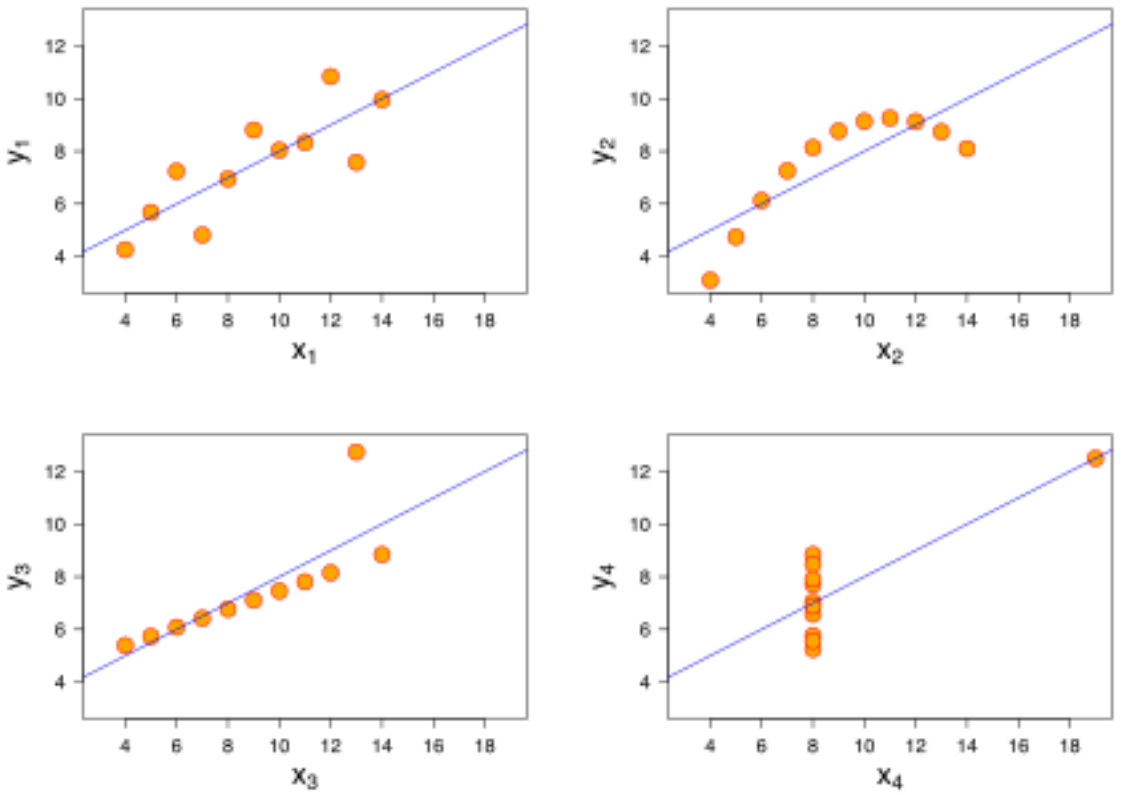


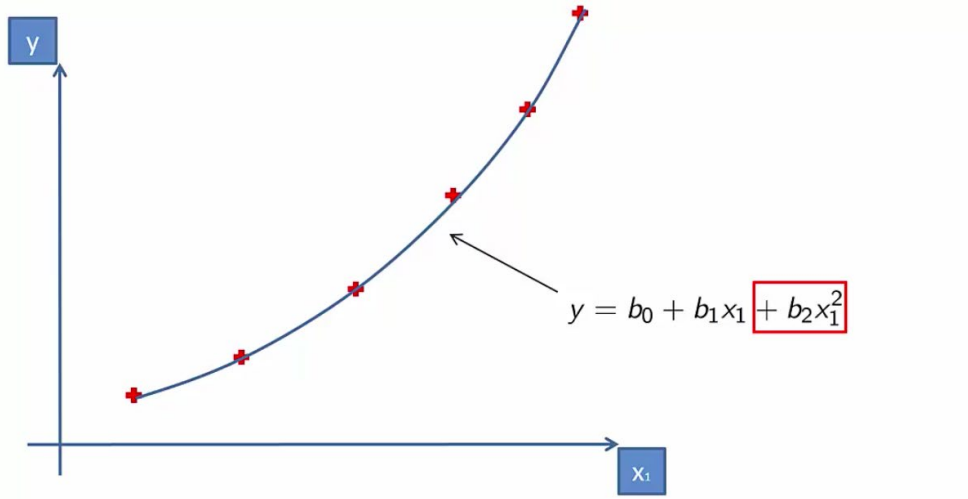


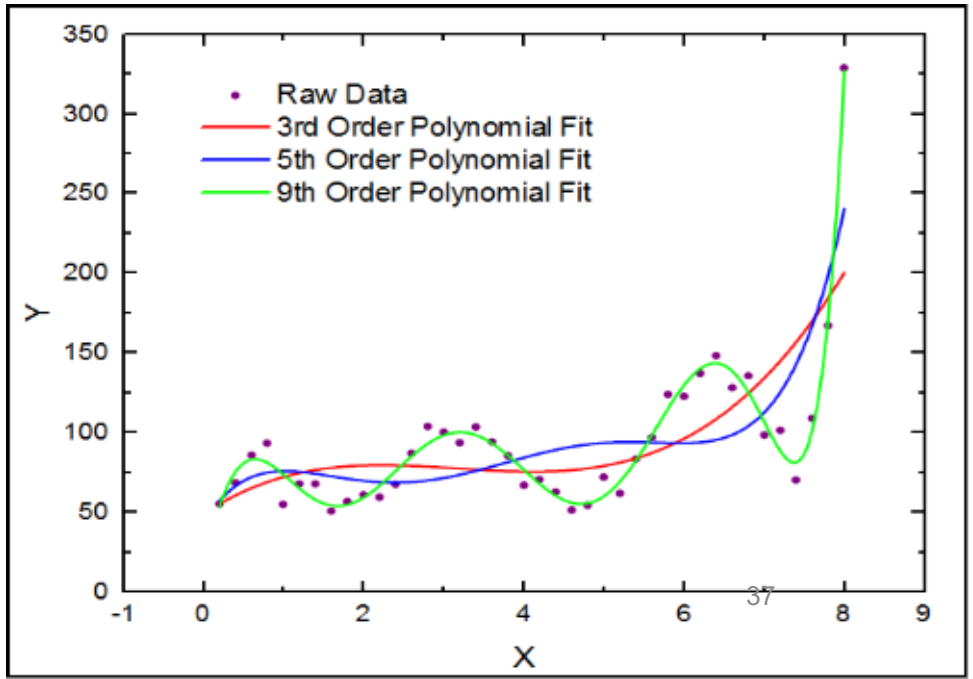
$$Y = a + bX$$

Find Y and X such that MSE is minimum.

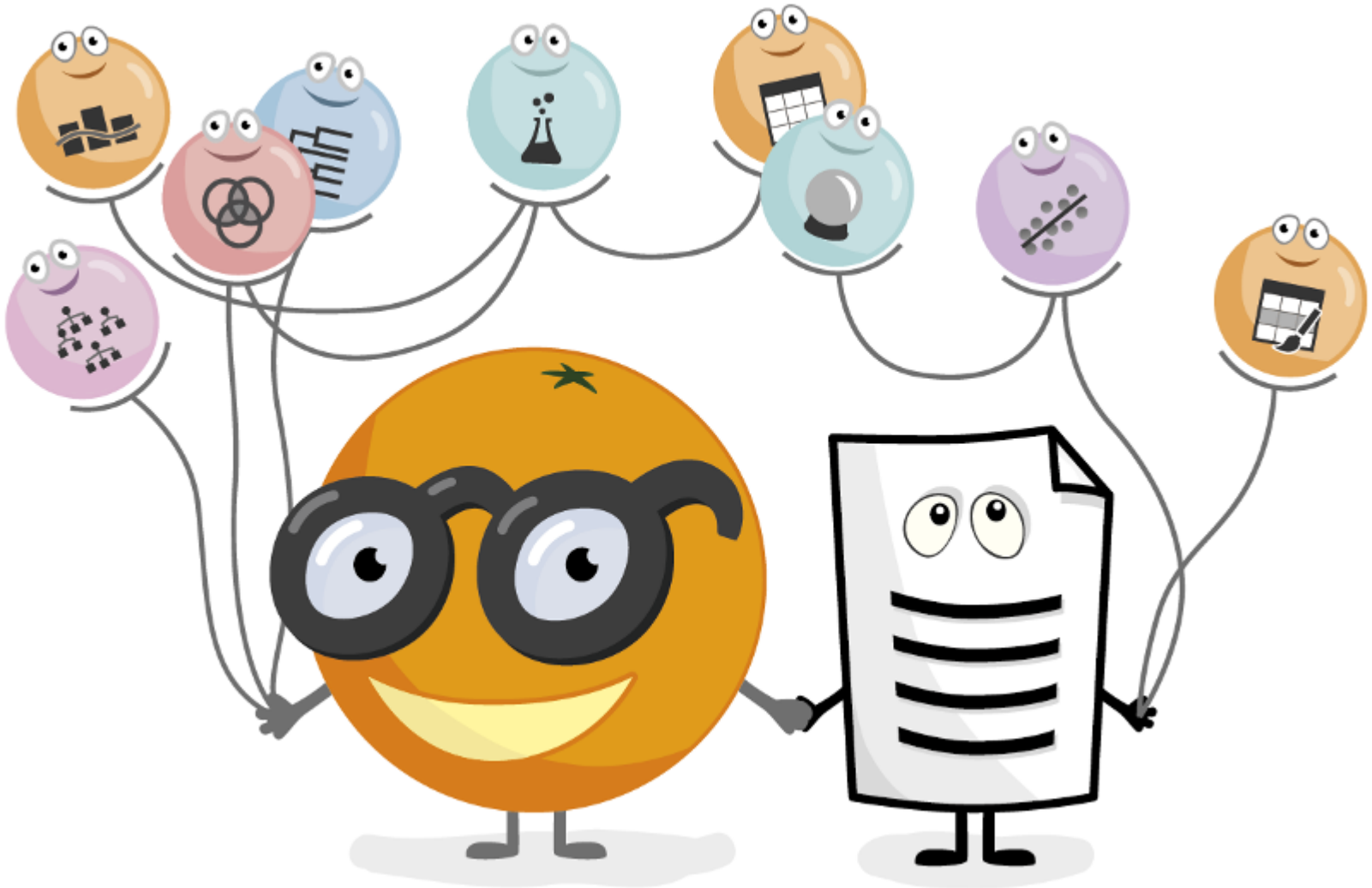
So you heard about linear regression, how about other, curvier lines…







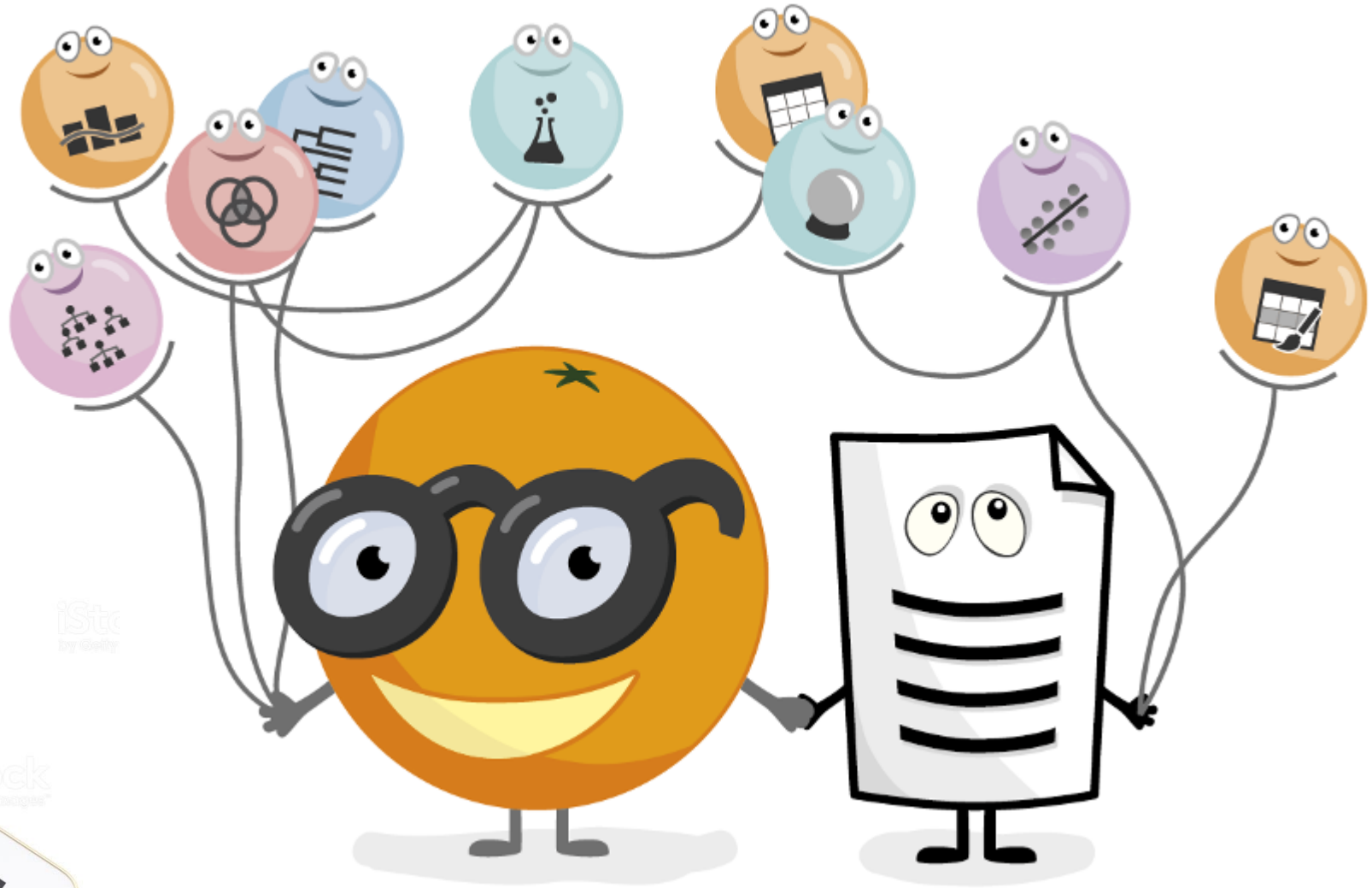
And now for some coffee, followed by some vitamins…

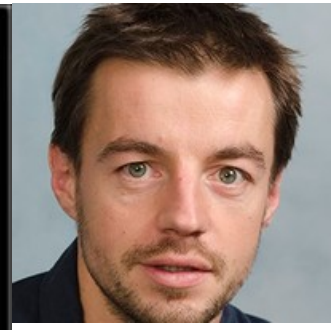


INTERMISSION 1

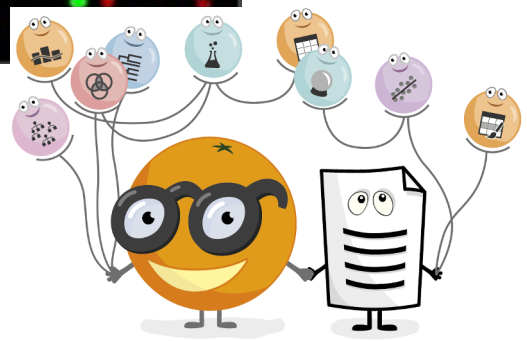
orange

Data Mining
Fruitful and Fun



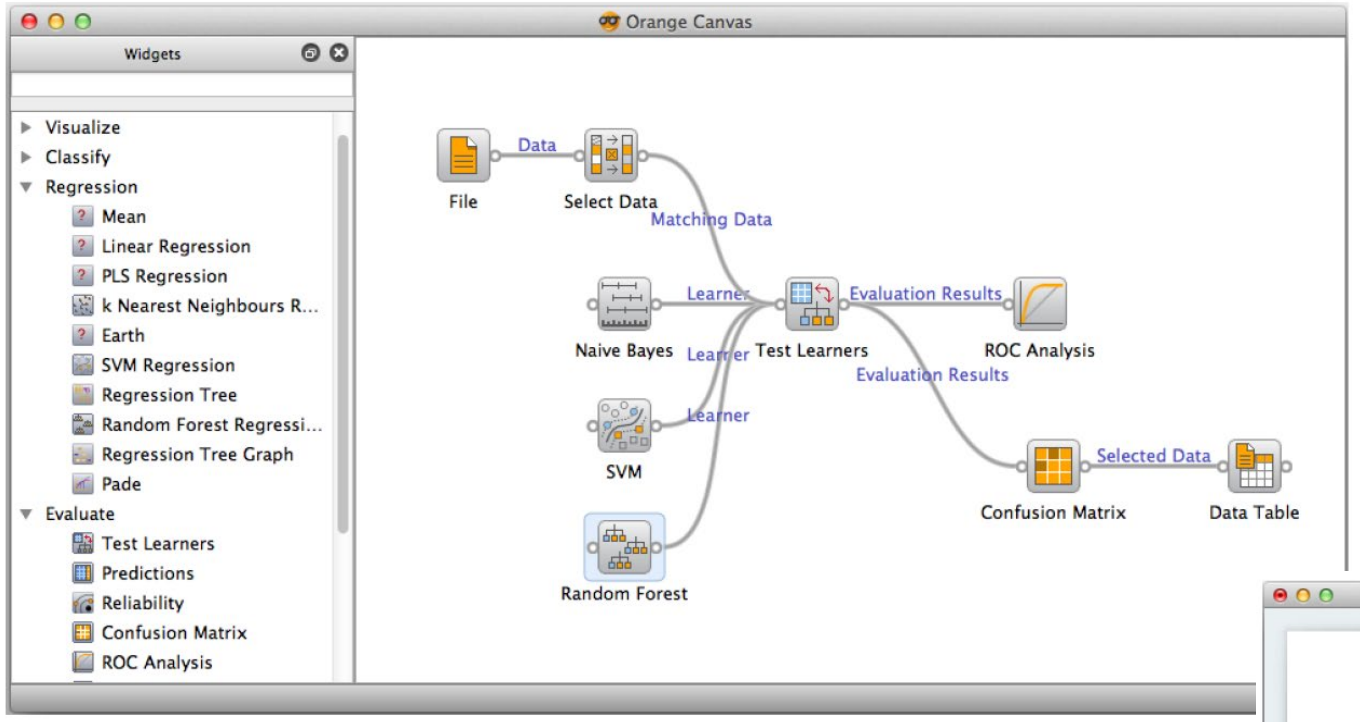


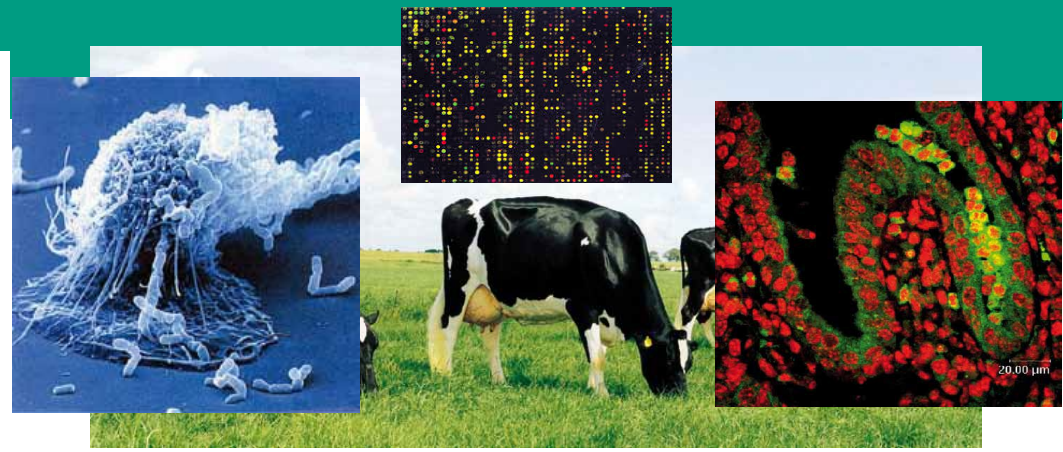
Informatica **37** (2013) 55–60



Orange: Data Mining Fruitful and Fun - A Historical Perspective

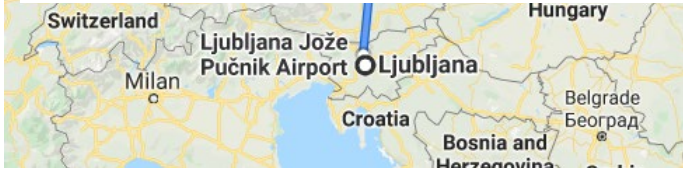
Janez Demšar and Blaž Zupan
University of Ljubljana, Faculty of Computer and Information Science, Tržaška 25, Ljubljana, Slovenia
E-mail: {janez.demsar|blaz.zupan}@fri.uni-lj.si

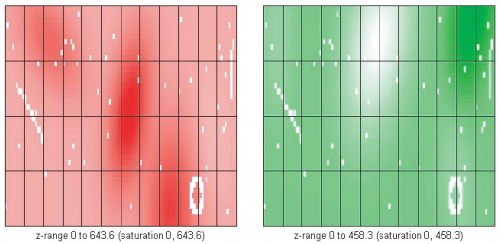




Genet. Sel. Evol. 39 (2007) 633–650
© INRA, EDP Sciences, 2007
DOI: 10.1051/gse:2007029

Available online at:
www.gse-journal.org

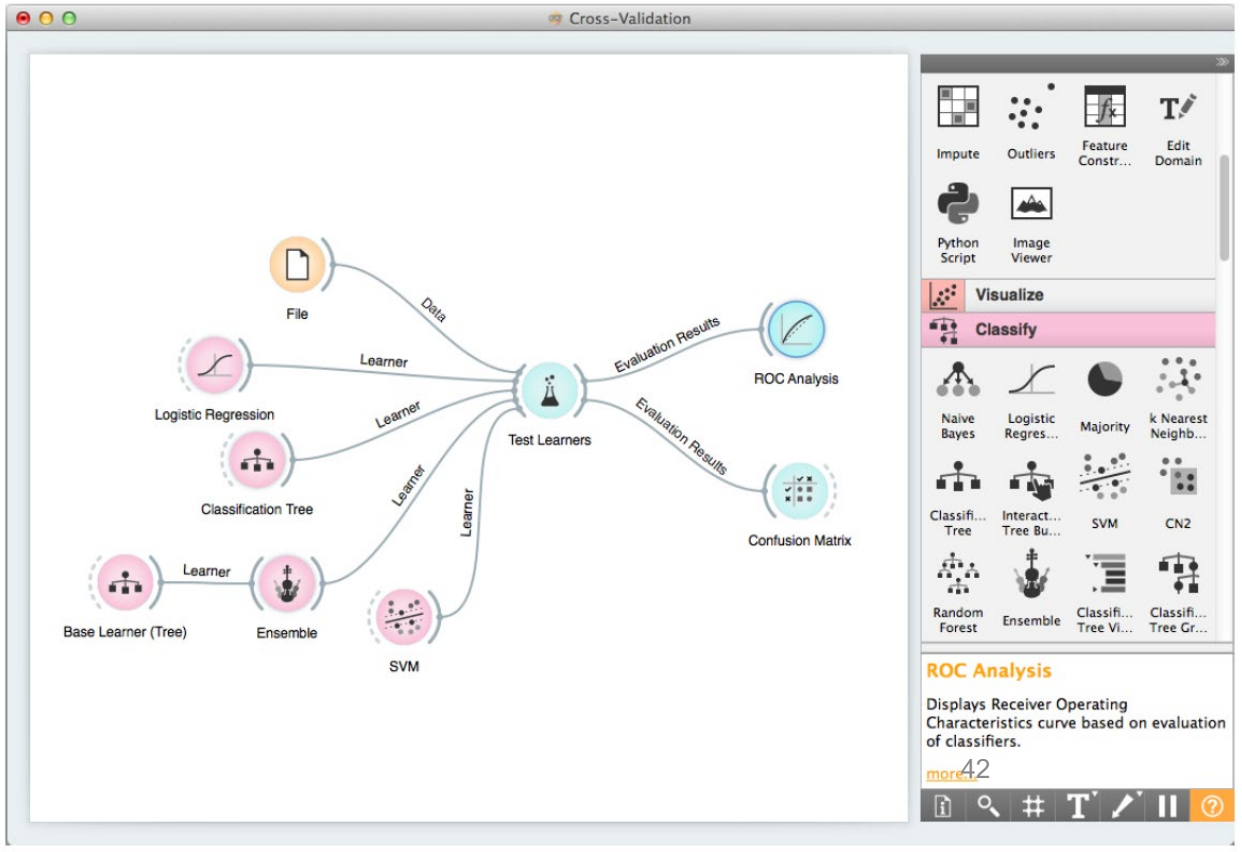


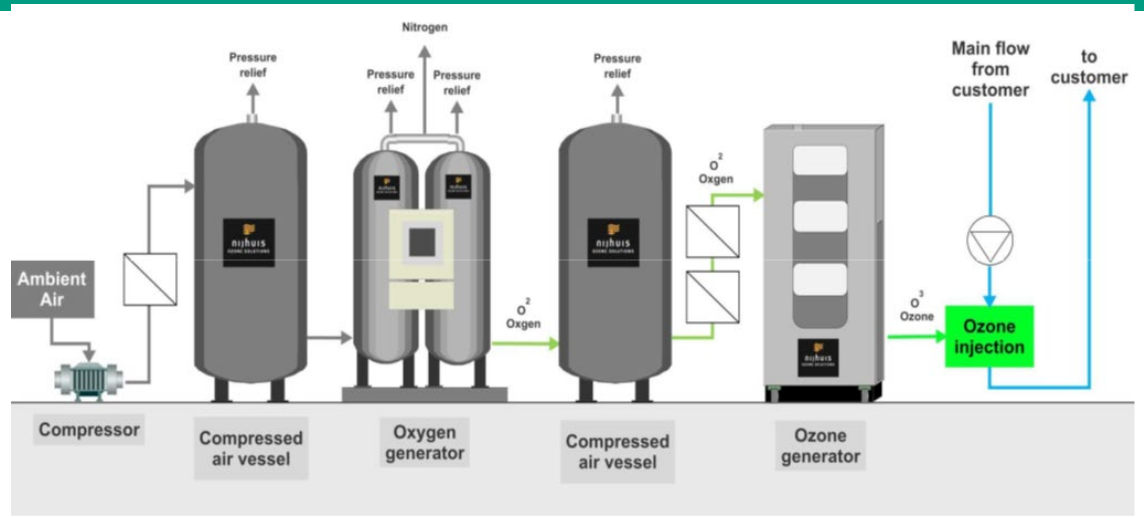


Genet. Sel. Evol. 39 (2007) 669–683
© INRA, EDP Sciences, 2007
DOI: 10.1051/gse:2007031

Analysis of a simulated microarra
Comparison of methods f
data normalisation and dete
of differential expressio
(*Open Access publication*

Michael WATSON[a*], Mónica PÉREZ-ALEGRE[b], M
BARON[c], Céline DELMAS[d], Peter DOVČ[e], Mylèr
Jean-Louis FOULLEY[f], Juan José GARRIDO-PA
HULSEGGE[g], Florence JAFFRÉZIC[f], Ángeles JIMÉ
Miha LAVRIČ[e], Kim-Anh LÊ CAO[h], Guillemette M
MOUZAKI[h], Marco H. POOL[c], Christèle ROBERT-GI
SAN CRISTOBAL[d], Gwenola TOSSER-KLOPP
WADDINGTON[h], Dirk-Jan DE KONING





Hanzehogeschool Groningen
University of Applied Sciences

SAXION

umcg

Roessingh
Centrum voor Revalidatie

EUROSPED
TopVolleybal in Twente

de Fysiotherapeut
Nederlandse Vereniging voor Fysiotherapie in de Sportgezondheidszorg

360 sports intelligence

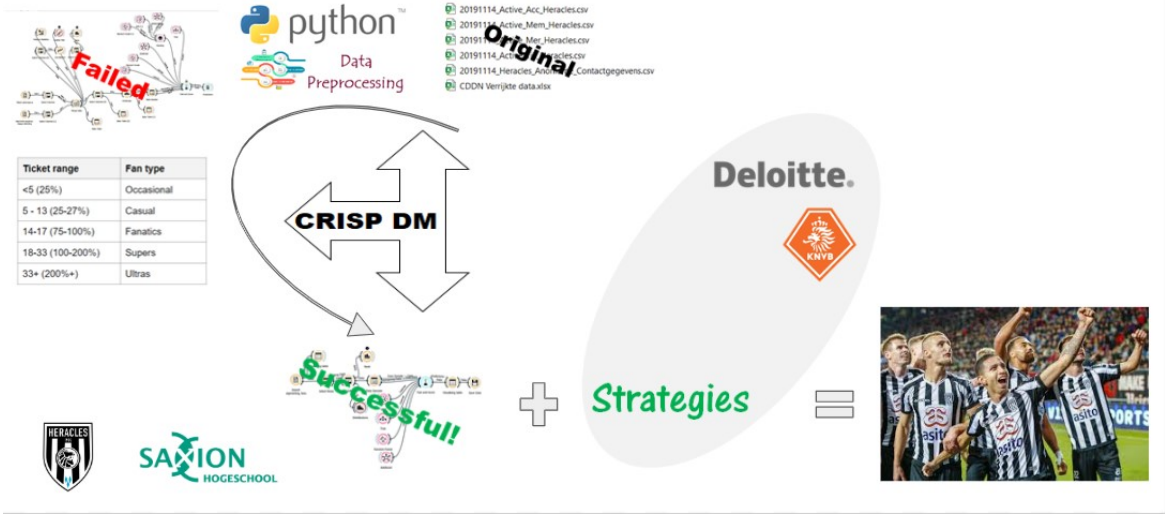
CE-Mate

FysioCentrum Nijverdal

TopvormTwente.

Big data technologie voor detectie overbelasting sporters

Visualisation



CastorFiber - labeled data - observed labels.ows*

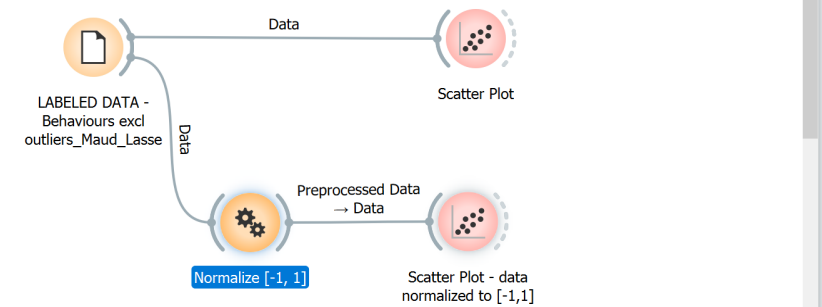
File Edit View Widget Options Help

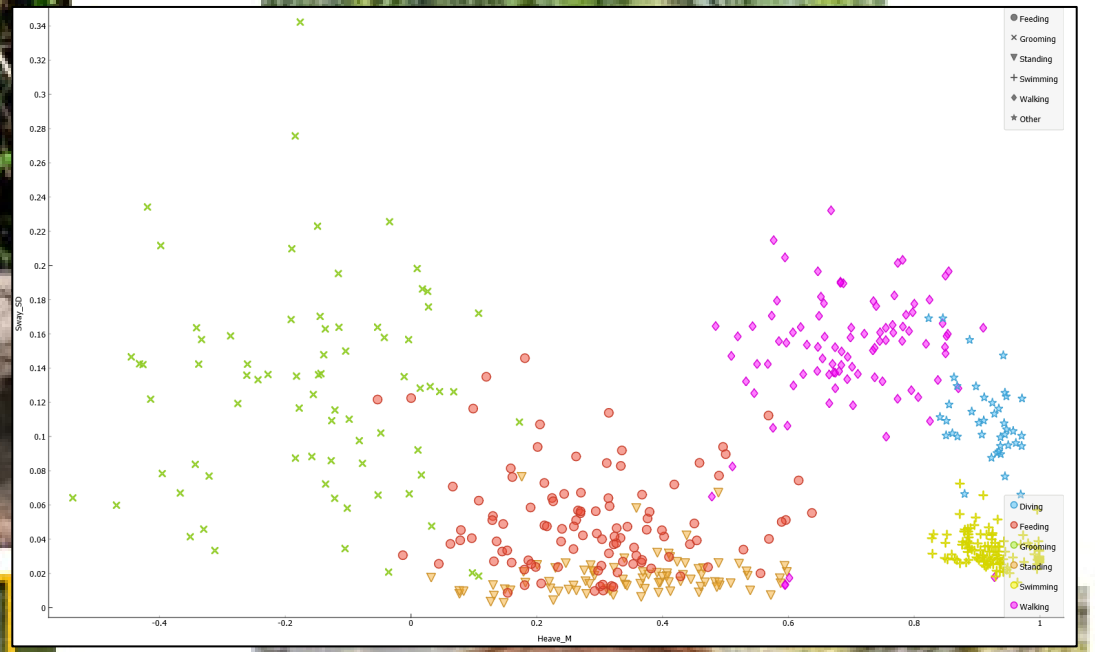
Data
Visualize
Model
Evaluate
Unsupervised
Prototypes
Image Analytics
Time Series
Text Mining
Single Cell

Preprocess

Construct a data preprocessing pipeline.

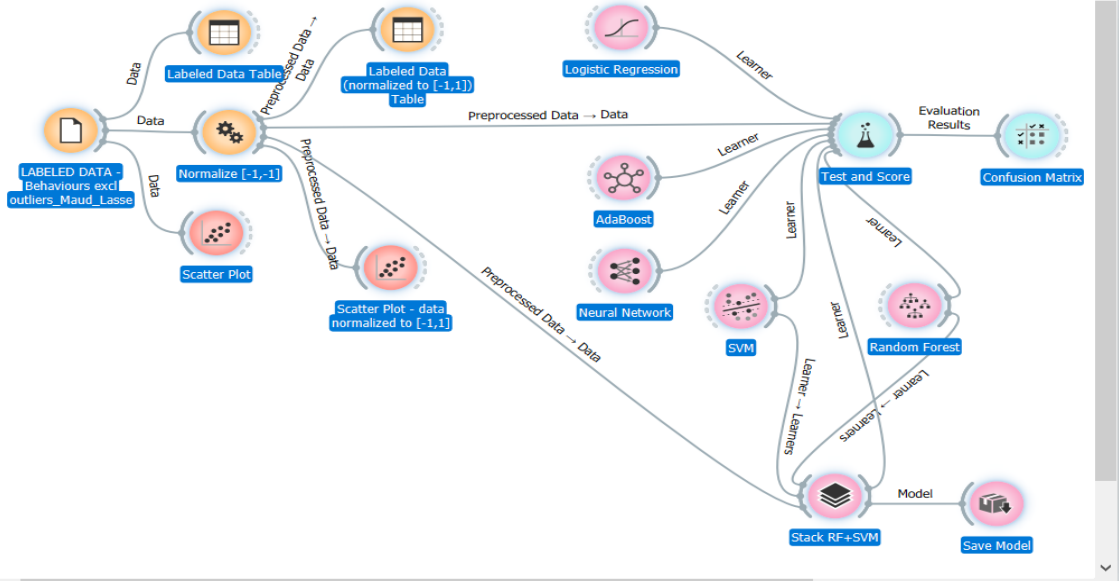
more...





CastorFiber - labeled data - testing machine learning models on observed labels data.ows*

File Edit View Widget Options Help



Test and Score - cross validation

Sampling

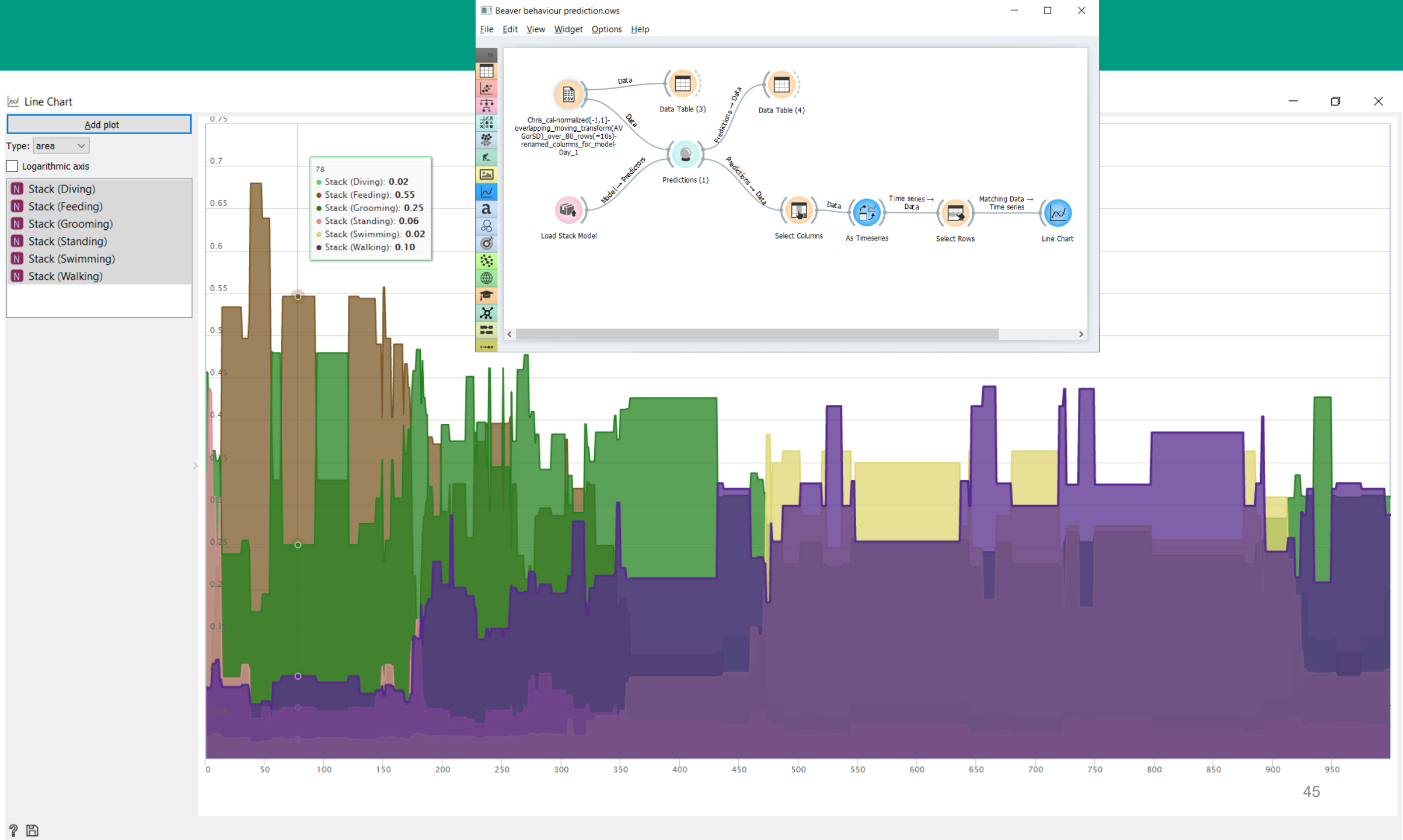
- Cross validation
  - Number of folds: 10
  - Stratified
- Cross validation by feature
- Random sampling
  - Repeat train/test: 10
  - Training set size: 80 %
  - Stratified
- Leave one out
- Test on train data
- Test on test data

Target Class

(Average over classes)

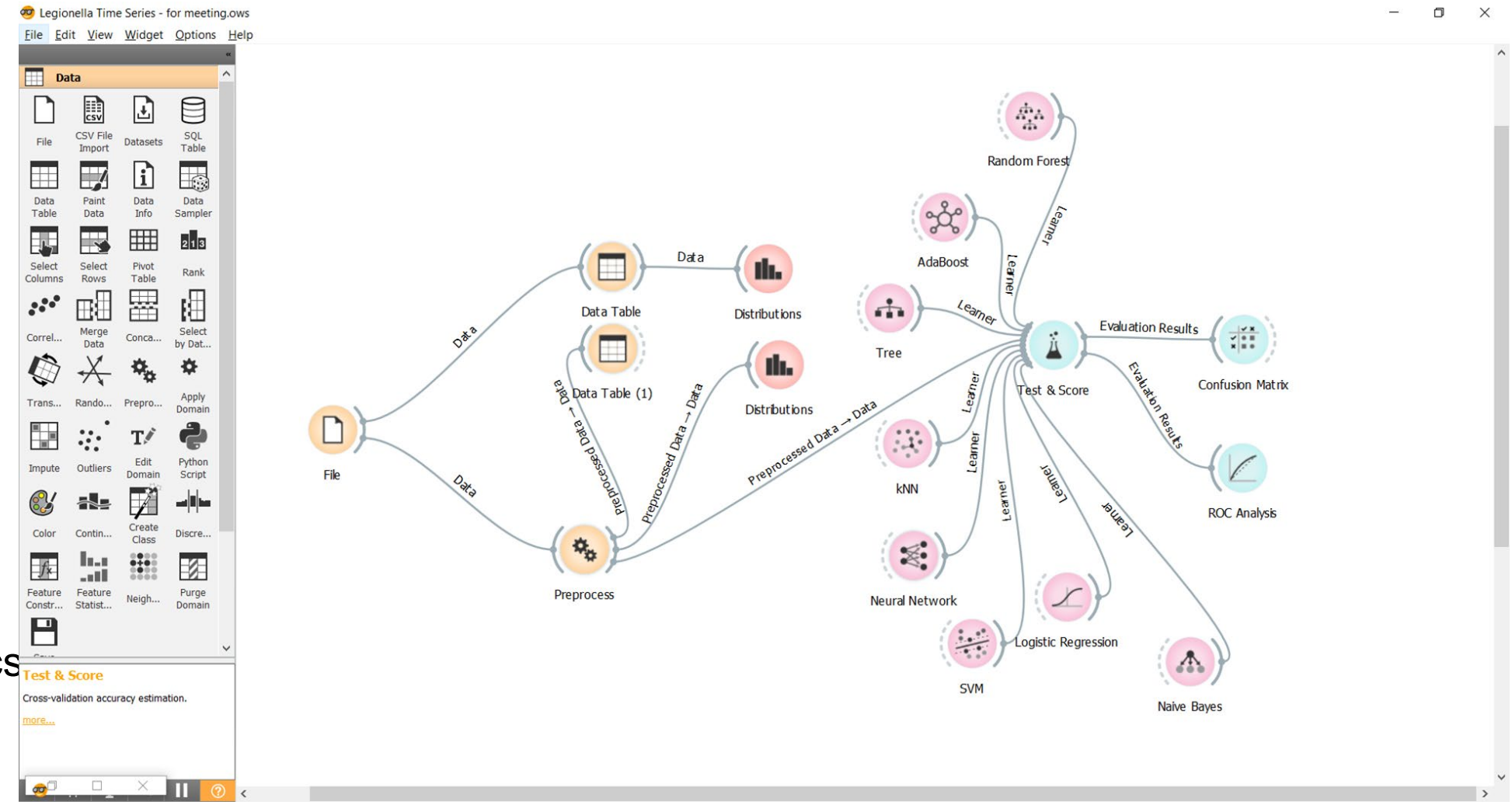
Evaluation Results

Model	AUC	CA	F1	Precision	Recall
Stack	0.992	0.938	0.938	0.939	0.938
SVM	0.990	0.930	0.931	0.939	0.930
Random Forest	0.988	0.924	0.925	0.925	0.924
kNN	0.988	0.913	0.913	0.921	0.913
Neural Network	0.987	0.922	0.923	0.926	0.922
Logistic Regression	0.983	0.899	0.899	0.904	0.899
AdaBoost	0.940	0.901	0.901	0.901	0.901


Beaver behaviour prediction.ows
File Edit View Widget Options Help
Data
Data Table (3)
Predictions → Data
Data Table (4)
Chris_cal-normalized[-1,1]-overlapping_moving_transform(AVGorSD)_over_80_rows(=10s)-renamed_columns_for_model-Day_1
Data
Predictions (1)
Model → Predictors
Predictors → Data
Load Stack Model
Select Columns
Data
As Timeseries
Time series → Data
Select Rows
Matching Data → Time series
Line Chart
Line Chart
Add plot
Type: area
Logarithmic axis
Stack (Diving)
Stack (Feeding)
Stack (Grooming)
Stack (Standing)
Stack (Swimming)
Stack (Walking)
78
Stack (Diving): 0.02
Stack (Feeding): 0.55
Stack (Grooming): 0.25
Stack (Standing): 0.06
Stack (Swimming): 0.02
Stack (Walking): 0.10
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
850
900
950

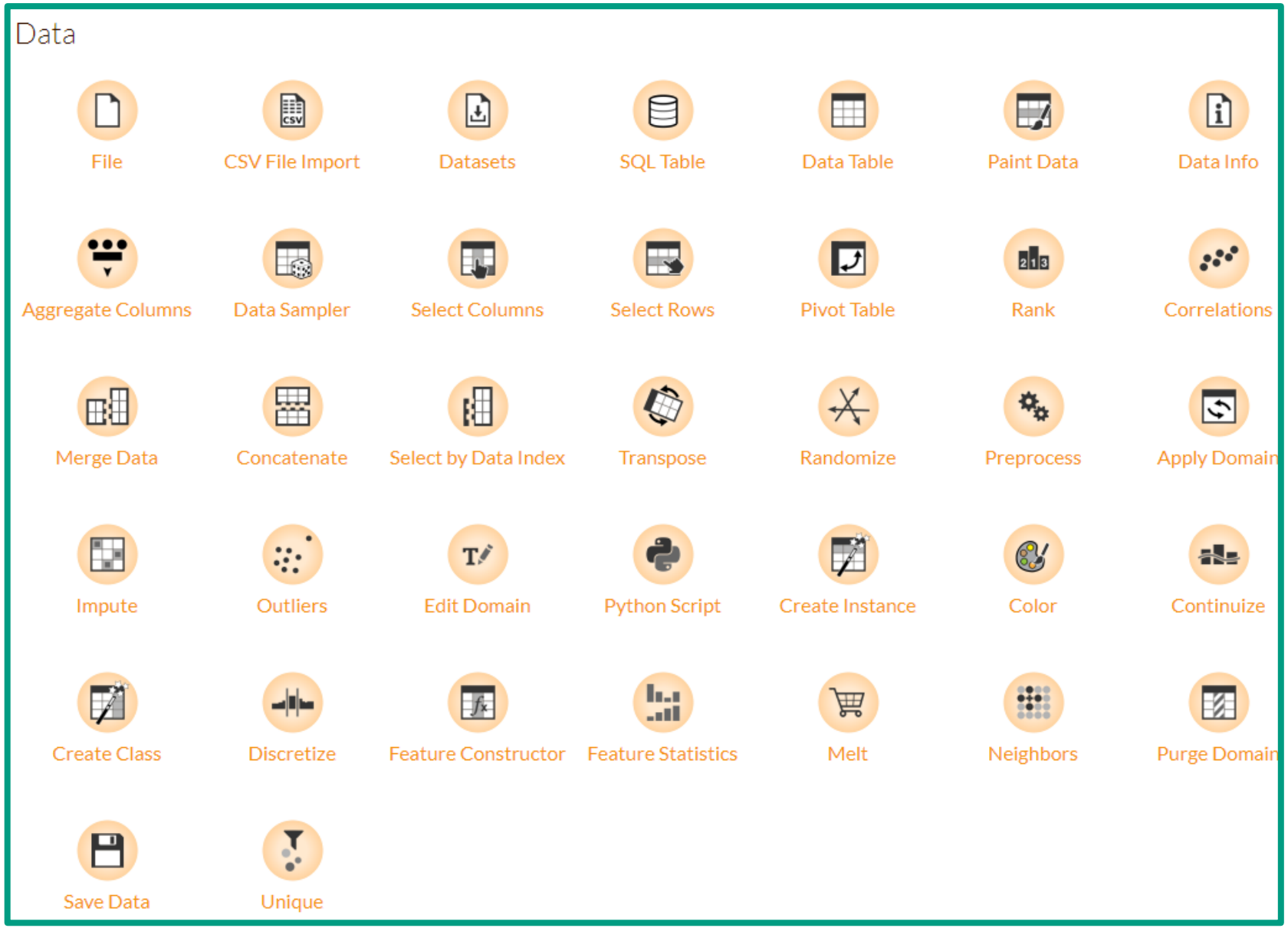
The Features of Orange

- **Canvas**: graphical front-end for data analysis (based on Python), also available in Anaconda
- **Widgets**:
  - Data
  - Visualize
  - Classify
  - Regression
  - Evaluate
  - Unsupervised
- **Add-ons:**
  - Associate
  - Bioinformatics
  - Data fusion
  - Educational
  - Geo
  - Image analytics
  - Network
  - Text mining
  - Time series
  - Spectroscopy

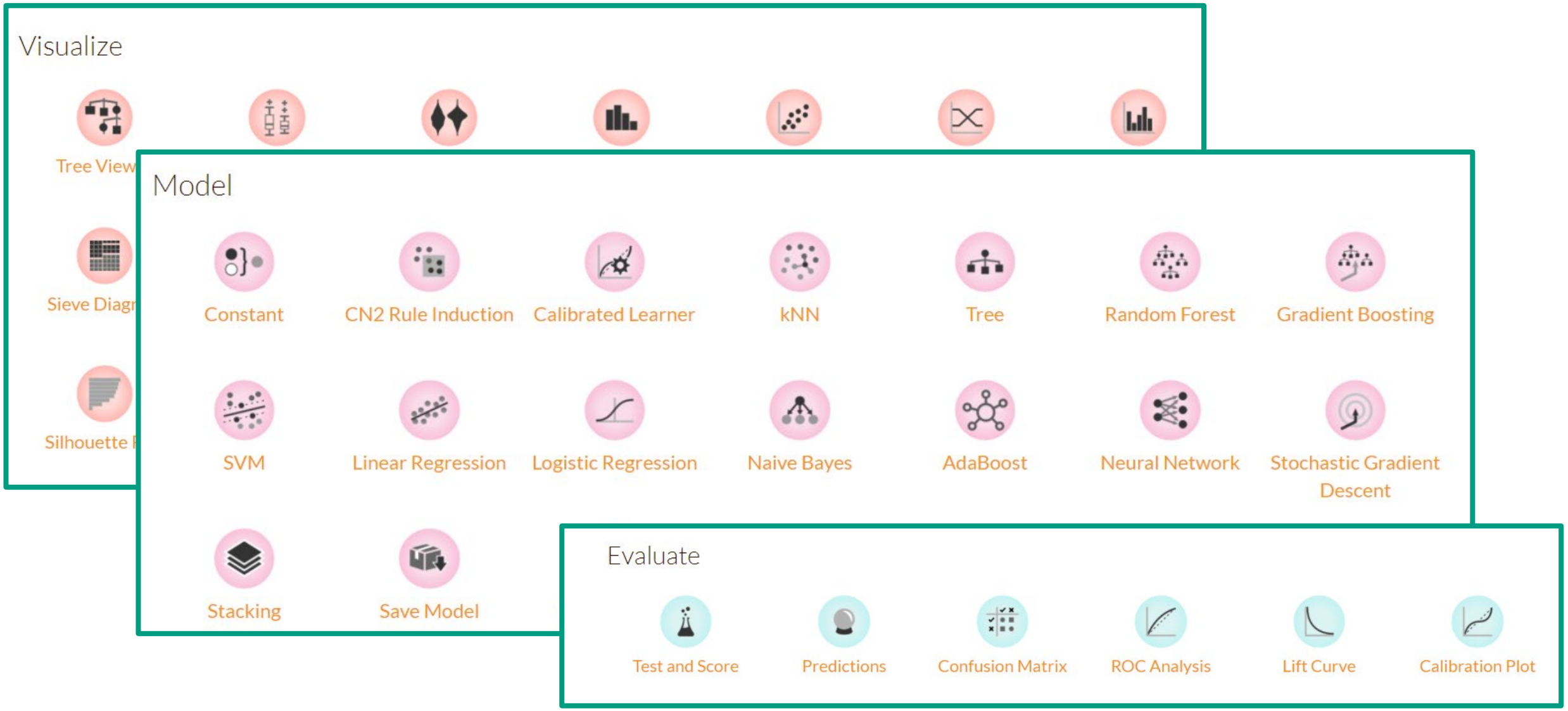


- Able to make a quick (sometimes dirty) machine learning analysis.

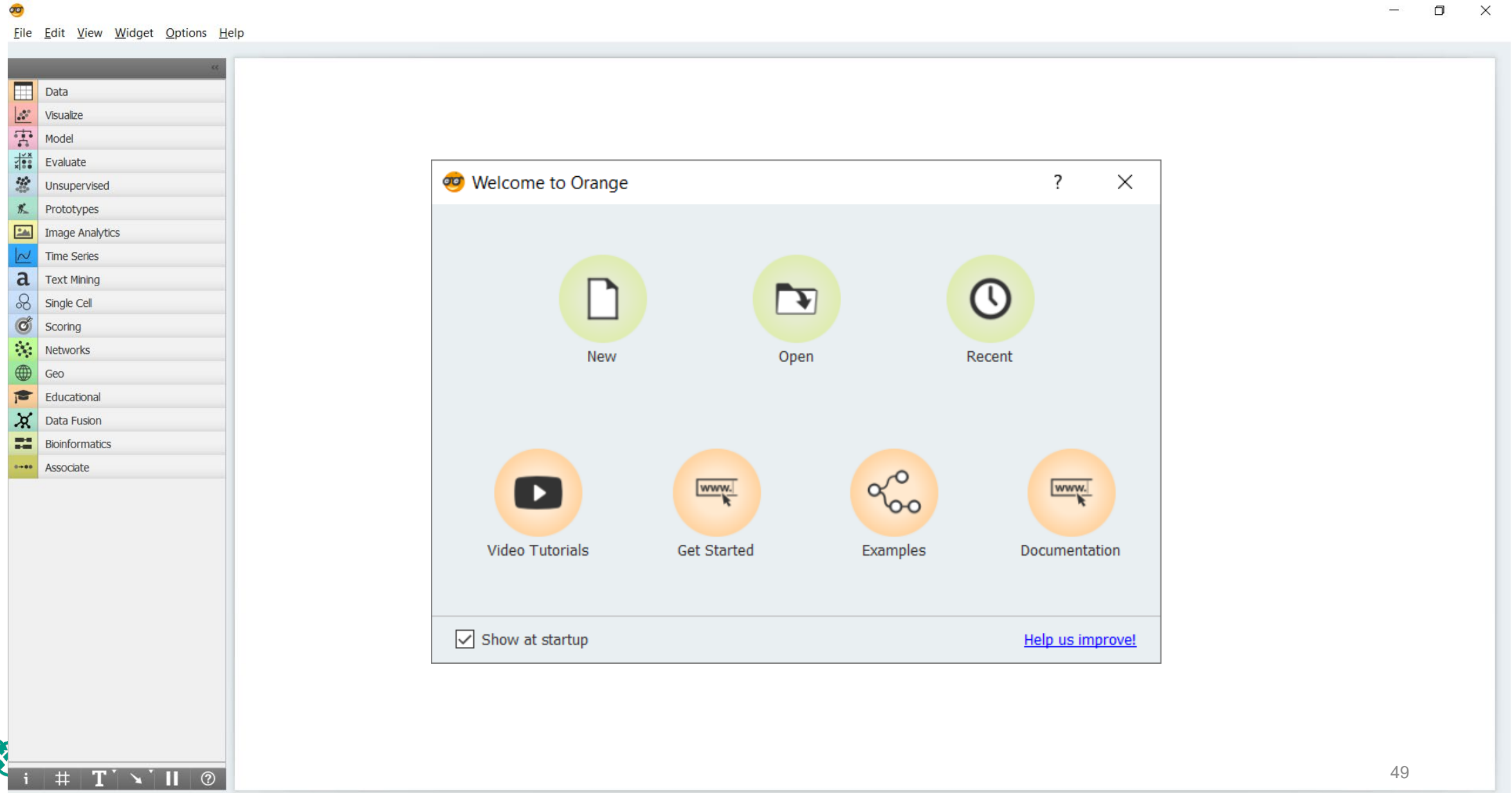
Features (more detailed) – Data manipulation



Features (more detailed) – Visualization, modelling, evaluation



And now for some Orange juice…



Additional resources

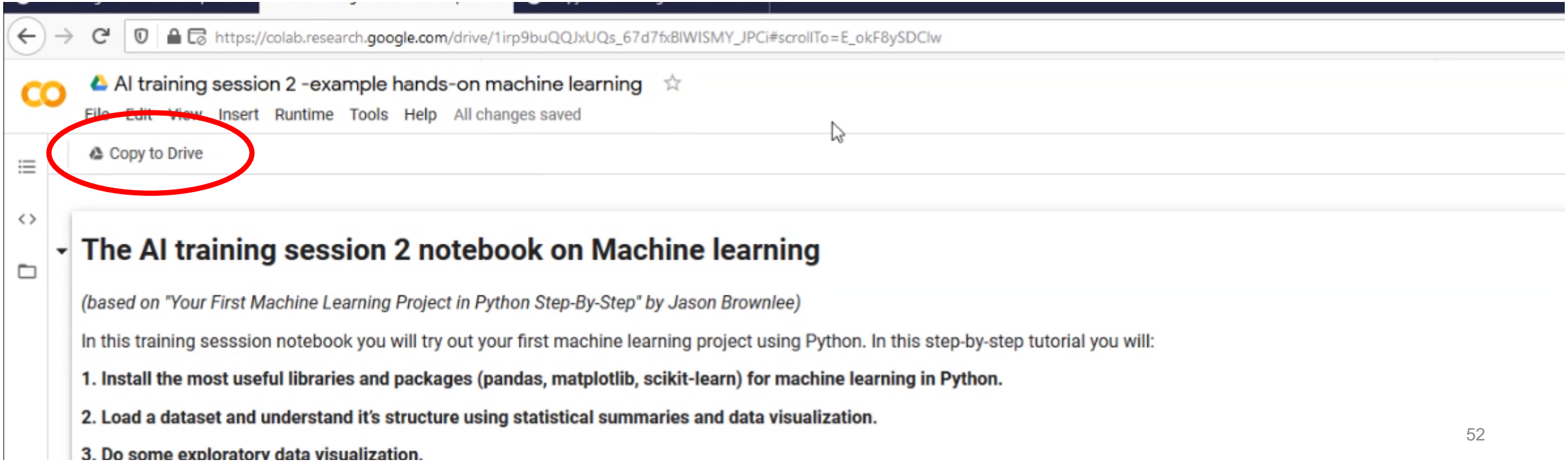
- Website - https://orange.biolab.si/
- Tutorials - https://www.youtube.com/playlist?list=PLmNPvQr9Tf-ZSDLwOzxpvY-HrE0yv-8Fy
- Stack Exchange - https://datascience.stackexchange.com/questions/tagged/orange
- Discord - https://discord.gg/FWrfeXV
- GitHub - https://github.com/biolab/orange3
- Facebook, Twitter

After the break, let us start with Hands-on ML in Colab….
The link will appear in the chat!

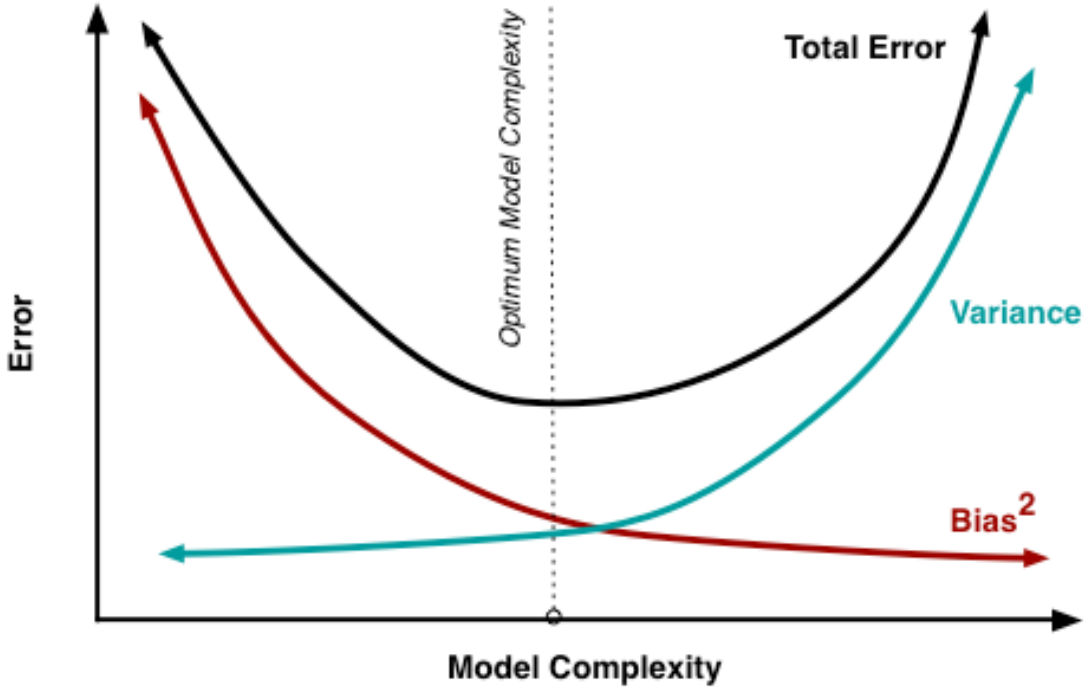
INTERMISSION 2

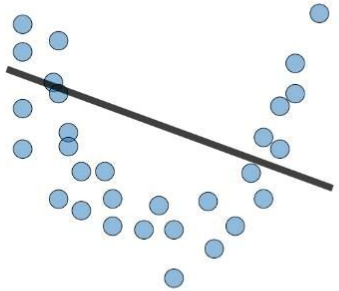
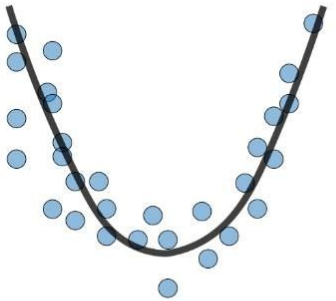
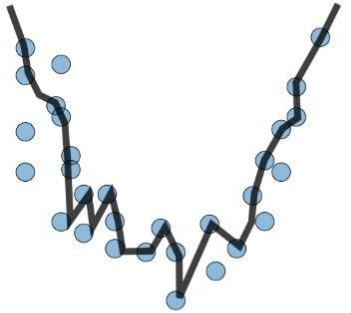
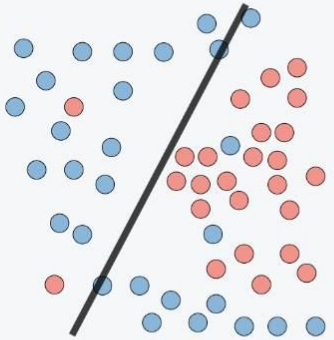
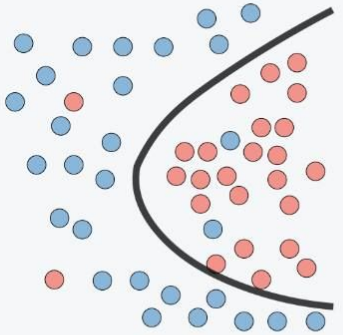
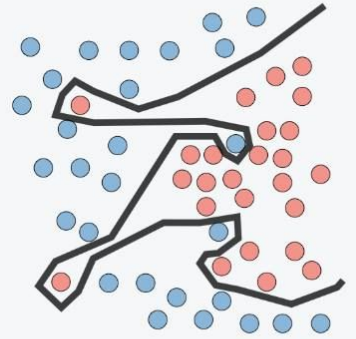
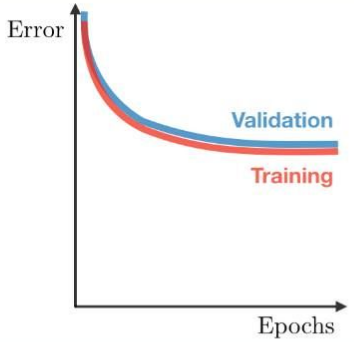
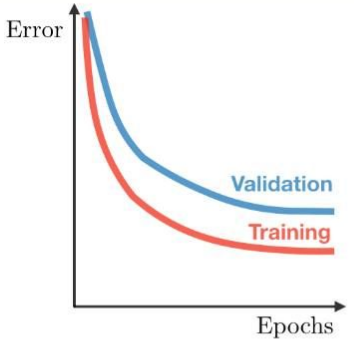
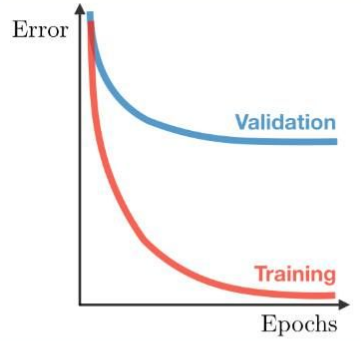
And now some hands-on in Colab with Iris

Link to the Python Colab Environment -
“AI training session 2 -example hands-on machine learning”
https://colab.research.google.com/drive/1irp9buQQJxUQs_67d7fxBIWISMY_JPCi
Be sure to push the button in the top left corner “Copy to Drive” so you can have your local copy and try things out! In the meantime, I will share my Colab screen…

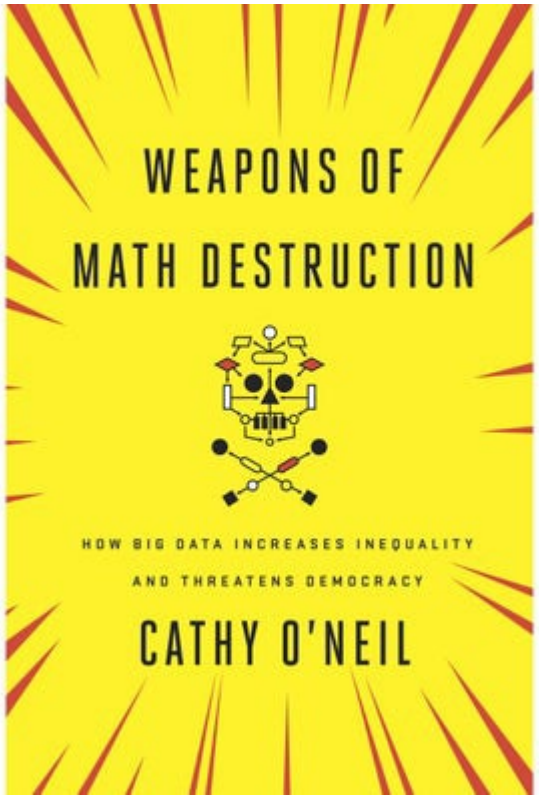


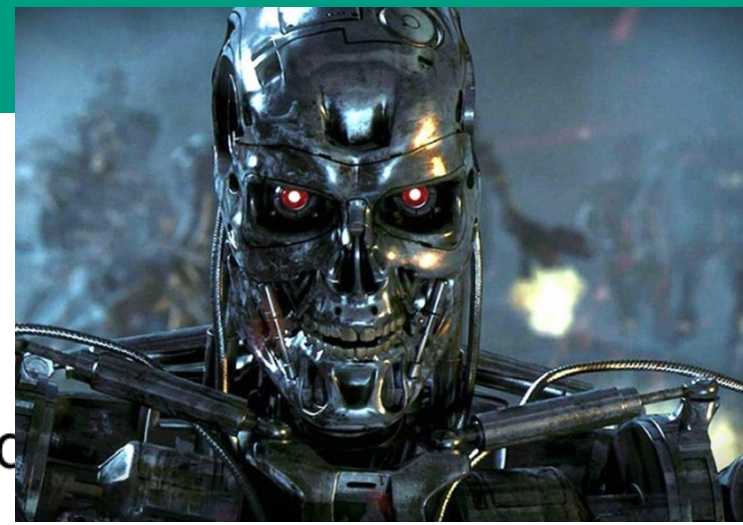
All these algorithms…
So, now what?
The bias-variance-complexity tradeoff!



	Underfitting	Just right	Overfitting
Symptoms	• High training error • Training error close to test error • High bias	• Training error slightly lower than test error	• Very low training error • Training error much lower than test error • High variance
Regression illustration			
Classification illustration			
Deep learning illustration			
Possible remedies	• Complexify model • Add more features • Train longer		• Perform regularization • Get more data

All these algorithms…
So, now what?
Accuracy versus explainability!





Accuracy

- Neural Networks
- Ensemble Methods (Ranc
- Support Vector Machine
- Graphical Models
- Decision Trees
- Regression Algorithms
- Classification Rules

Explainability

Recap

Learning goals

- Get to know machine learning:
  - Principles
  - Algorithms
  - Evaluation

- Hands-on experience in ML
  - Visual programming – Orange
  - Coding – Python in Colab

Open questions?

All learning materials will be made available.

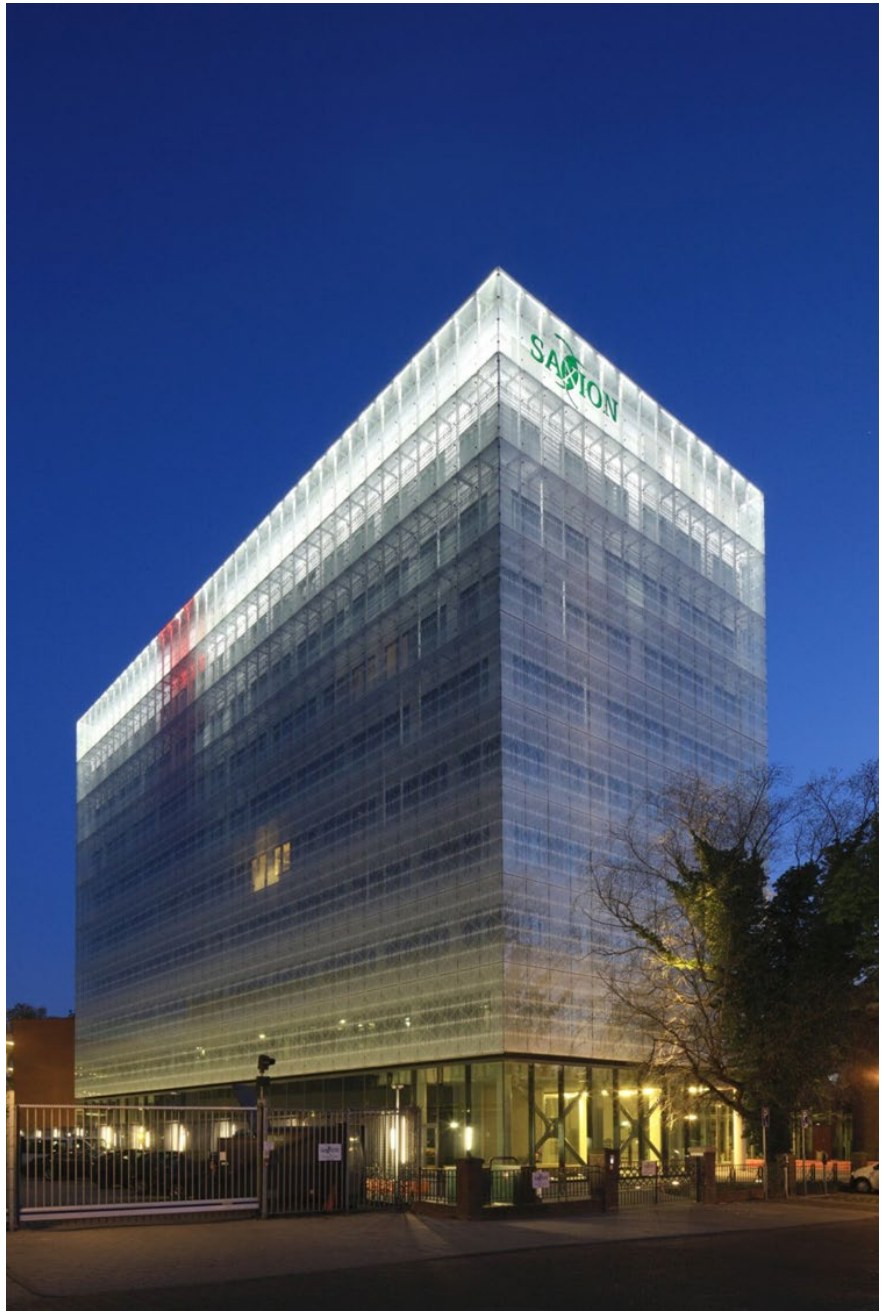
Thanks for your attention!

Next session:

18 June:
Deep learning

Send an email!

m.lavric@saxion.nl



Visit our websites!

saxion.nl/ami

tvalley.nl

boostsmartindustry.nl

SAXION
UNIVERSITY OF APPLIED SCIENCES

Media sources 1

Slovenia: http://countriestips.over-blog.com/2020/04/what-is-slovenia-known-for.html
University of Ljubljana: www.uni-lj.si
DSMBS: https://dsmbs.wordpress.com/
Ljubljana: https://www.nomadepicureans.com/
Pagellus erythrinus: https://www.fishbase.se/photos/PicturesSummary.php?ID=893&what=species
Red Star Chicken: https://www.flickr.com/photos/33149580@N08/3085186294/
Microarray: https://biology4alevel.blogspot.com/2016/07/158-genetic-markers-and-microarrays.html
Orange: https://orange.biolab.si/
Gluten Free Icon: https://www.pngitem.com/middle/hJiibTh_gluten-free-icon-vector-hd-png-download/
Trieste: https://ouritaliantable.com/trieste-goulash-goulash-triestino/
Trieste: https://www.inntravel.co.uk/city-add-ons/trieste
University of Trieste: https://www.units.it
Schematic representation of the TG2 interactome: Altuntas et al. The transglutaminase type 2 and pyruvate kinase isoenzyme M2 interplay in autophagy regulation. Oncotarget. 2015; 6: 44941-44954.
University of Muenster: https://www.uni-muenster.de
University Hospital Muenster: https://www.ukm.de/
Münster location on the Germany map: http://ontheworldmap.com/germany/city/munster/munster-location-on-the-germany-map.html
Vector illustration of intestine with Crohn's disease: https://www.dreamstime.com/vector-medical-illustration-section-normal-intestine-compared-to-symptoms-crohn-s-disease-crohns-image119957874
Kawasaki disease - https://www.aboutkidshealth.ca/fr/Article?contentid=915&language=English
Contribution of alarmins to certain stages of inflammatory arthritis, (A) Preactivation of innate immune cells by a complex set of factors: Lavric et al. Alarmins firing arthritis: Helpful diagnostic tools and promising therapeutic targets. Joint Bone Spine. 2017;84(4):401-410.
Visual output of the LIME sensitivity analysis for end users: Amrit et al. Identifying child abuse through text mining and machine learning. Expert systems with applications. 2017 Dec 1;88:402-18. MRDM: https://mrdm.nl
Nomogram: https://www.utwente.nl/en/techmed/influence/nomogram/
HFGP: http://www.humanfunctionalgenomics.org/site/
Neural network algorithm: Alaoui et al. "Data Mining and Machine Learning Approaches and Technologies for Diagnosing Diabetes in Women." International Conference on Big Data and Networks Technologies. Springer, Cham, 2019.

Media sources 2

Beaver: Jeff R Clow / Getty Images
Nijhuis industries ozone generation
Magic Crystal Ball Clipart: https://www.pngitem.com/middle/hbmbbRh_magic-crystal-ball-clipart-transparent-cartoons-magic-crystal/
Cartoon Wizzard: https://www.gograph.com/clipart/cartoon-wizard-gg67374029.html
What's the difference between #datascience, #machinelearning, and #artificialintelligence? by @drob https://twitter.com/drob
“No, Machine Learning is not just glorified Statistics”- https://towardsdatascience.com/no-machine-learning-is-not-just-glorified-statistics-26d3952234e3 original comic by sandserif https://www.instagram.com/sandserifcomics/
An Introduction to Statistical Learning: https://link.springer.com/book/10.1007/978-1-4614-7138-7
The Elements of Statistical Learning: https://link.springer.com/book/10.1007/978-0-387-84858-7
Analysis of gene ranking by classical inference and ML: Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. Nat Methods 15, 233–234 (2018).
Machine Learning (Dilbert by Scott Adams): https://dilbert.com/strip/2013-02-02
AI Machine Learning Data Visualization: Image Result For Ai Machine Learning Data Visualization https://quantumcomputingtech.blogspot.com/2019/01/machine-learning-data-visualization.html
A Machine Learning Workflow by Frederick Giasson: https://quantumcomputingtech.blogspot.com/2019/09/supervised-machine-learning-workflow.html
Data Scientists Spend Most of Their Time Cleaning Data: https://whatsthebigdata.com/2016/05/01/data-scientists-spend-most-of-their-time-cleaning-data/
Percentage of Time Allocated to Machine Learning Project Tasks: https://www.cloudfactory.com/data-labeling-guide
Visual Representation of Train/Test Split and Cross Validation: Author: Joseph Nelson, from Train/Test Split and Cross Validation in Python https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6
How do I evaluate a model? : https://www.quora.com/How-do-I-evaluate-a-model
Supervised learning model: Talwar and Yogesh. Machine Learning: An artificial intelligence methodology. International Journal of Engineering and Computer Science 2 (12) 3400-3404 (2013)
Regression vs Classification: https://towardsdatascience.com/regression-or-classification-linear-or-logistic-f093e8757b9c
Mickey Mouse, Donald Duck, Daisy Duck: Walt Disney Productions
Bugs Bunny, Daffy Duck: Warner Bros.
Precision and Recall: https://en.wikipedia.org/wiki/Precision_and_recall
DECISION TREES: https://lecturenotes.in/notes/26695-note-for-aerial-remote-sensing-and-photogrammetry-arsp-by-saneesh-ps?reading=true
Decision trees: Daniel T. Larose and Chantal D. Larose. Data Mining and Predictive Analytics, 2nd Edition, John Wiley & Sons, 2015

Media sources 3

Random forest: Nicolas Spies, Washington University, 2015.
Welcome to the forest, here's a random tree: https://medium.com/@saketh.ramanujam98/the-amazing-story-of-random-forests-2a3c3ef05b5d
Naïve Bayes classification: Rod Pierce et al. MathIsFun, 2014.
Support vector machines: Matthew Kelly, Computer Science: Source, 2010
A man feeding swans in the snow: By Marcin Ryczek from https://towardsdatascience.com/linear-discriminant-analysis-lda-101-using-r-6a97217a55a6
LDA: https://sebastianraschka.com/Articles/2014_python_lda.html
K-Nearest Neighbors (KNN) Algorithm: https://towardsdatascience.com/k-nearest-neighbors-knn-algorithm-bd375d14eec7
K-Nearest Neighbor Algorithm: https://towardsdatascience.com/implement-k-nearest-neighbors-classification-algorithm-c99be8f14052
Made in Slovenia: https://www.123rf.com/photo_96306822_made-in-slovenia-concept-3d-rendering-isolated-on-white-background.html
Orange: https://orange.biolab.si/
EADGENE: https://www.fabretp.eu/uploads/2/3/1/3/23133976/id_effab_brochurea5_digi2.pdf
CFGBC: http://cfgbc.mf.uni-lj.si/
Orange Schema old and new: Demšar & Zupan. Orange: Data mining fruitful and fun-a historical perspective. Informatica. 2013;37(1).
Example background plots: Watson et al. Analysis of a simulated microarray dataset: Comparison of methods for data normalisation and detection of differential expression. Genetics Selection Evolution 39(6) (2007): 669.
How to calculate MSE criteria in RandomForestRegression?: https://stackoverflow.com/questions/56369519/how-to-calculate-mse-criteria-in-randomforestregression
R-squared: https://www.rapidinsight.com/wp-content/uploads/r-squared.png
Fitting data to curve: Anscombe's Quartet in visual form (Anscombe, 1973).
Simple Linear Regression, Polynomial Regression: https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec
Relationship of bias, variance, generalization error and model complexity: Tan, Pang-Ning, M Steinbach and V Kumar, Introduction to data mining, 1, Boston: Pearson Addison Wesley, 145-195, 2006.
Underfitting, Just right, Overfitting: https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-machine-learning-tips-and-tricks#classification-metrics
Weapons of Math Destruction: Cathy O'Neil. Broadway Books. 2016
Terminator: James Cameron, Gale Anne Hurd. Paramount Pictures, 20th Century Fox.
Spicnik and the trip into the heart of vineyards: https://www.nina-potuje.com/kam-na-izlet-spicnik-pot-srce/