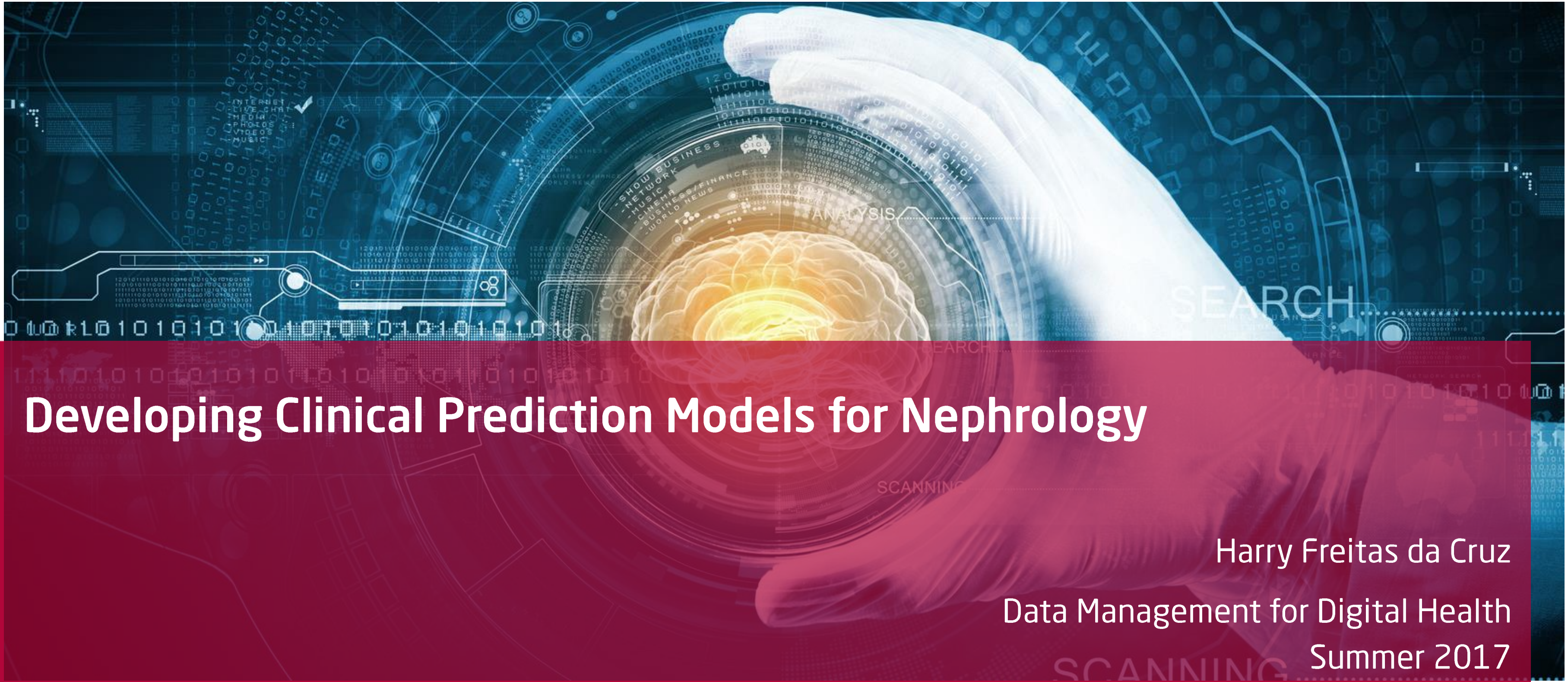


Developing Clinical Prediction Models for Nephrology

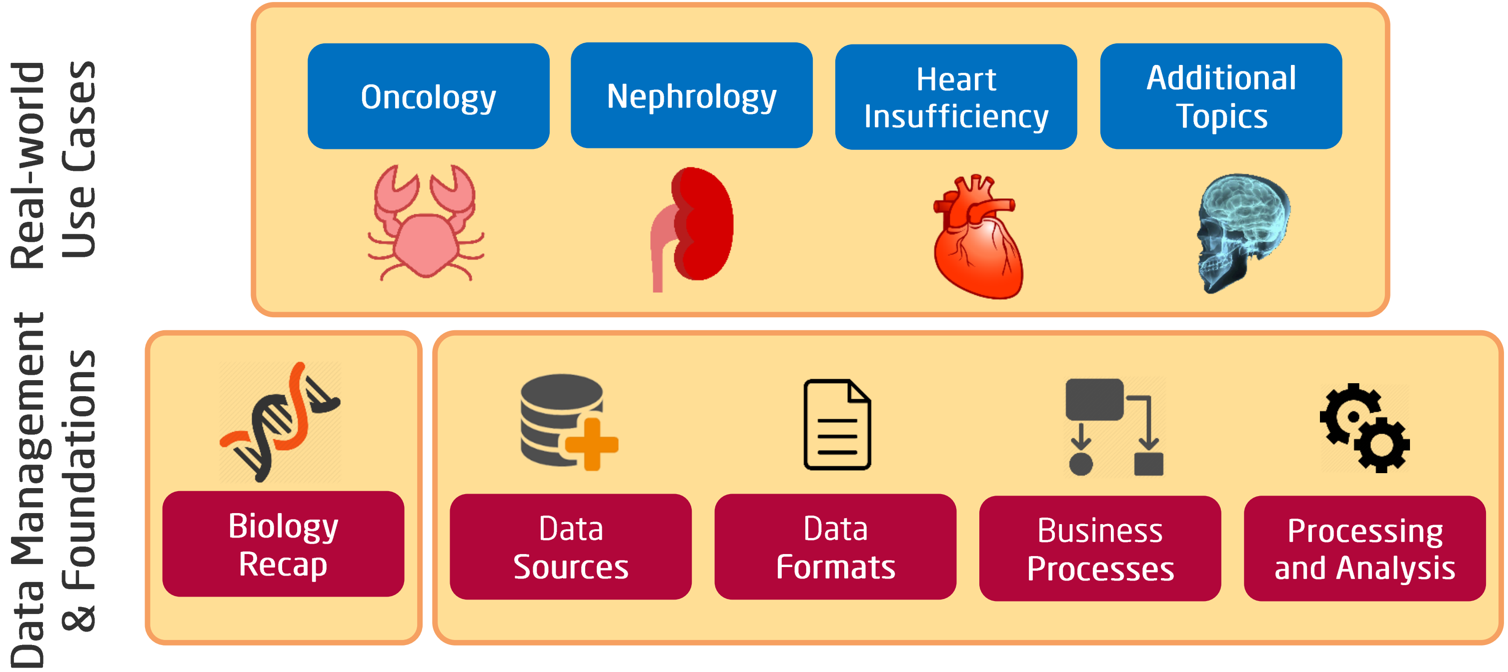
Harry Freitas da Cruz

Data Management for Digital Health

Summer 2017

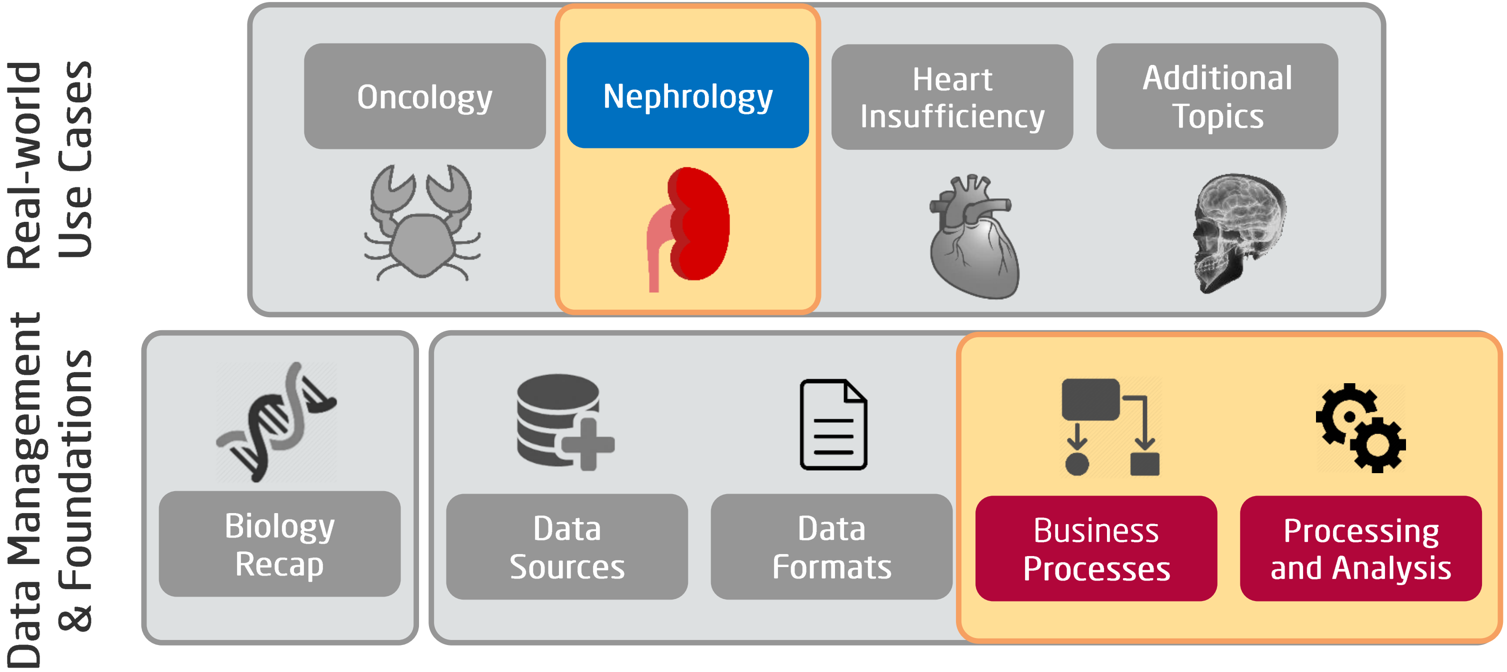
Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Where are we?



Real-world Use Cases
Oncology
Nephrology
Heart Insufficiency
Additional Topics
Data Management & Foundations
Biology Recap
Data Sources
Data Formats
Business Processes
Processing and Analysis

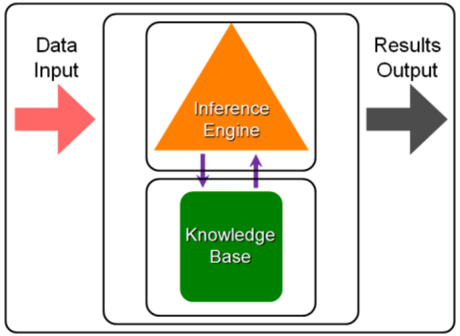
Where are we?



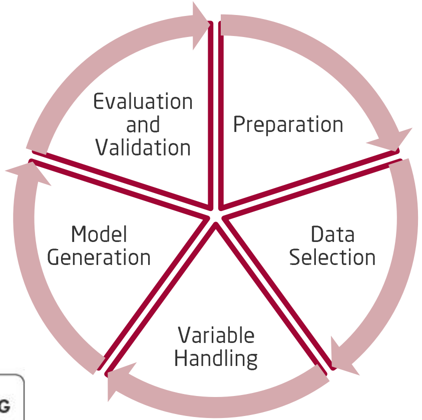
Real-world Use Cases
Oncology
Nephrology
Heart Insufficiency
Additional Topics
Data Management & Foundations
Biology Recap
Data Sources
Data Formats
Business Processes
Processing and Analysis

Recap Predictive Modelling for Clinical Applications

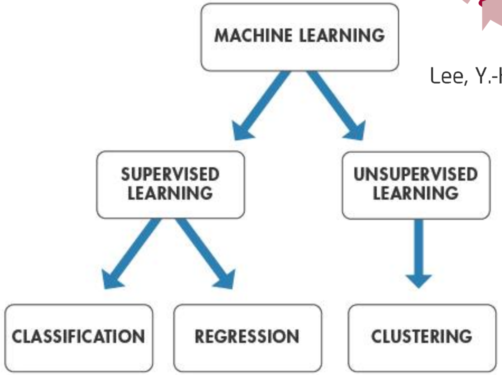
- Predictive analytics in healthcare
- Clinical Decision Support Systems
- Clinical Data Repository
- Developing a Clinical Prediction Model
  - Preparation
  - Data selection
  - Variable handling
  - Model generation



Architecture components of CDSS (Kola, n.d.)



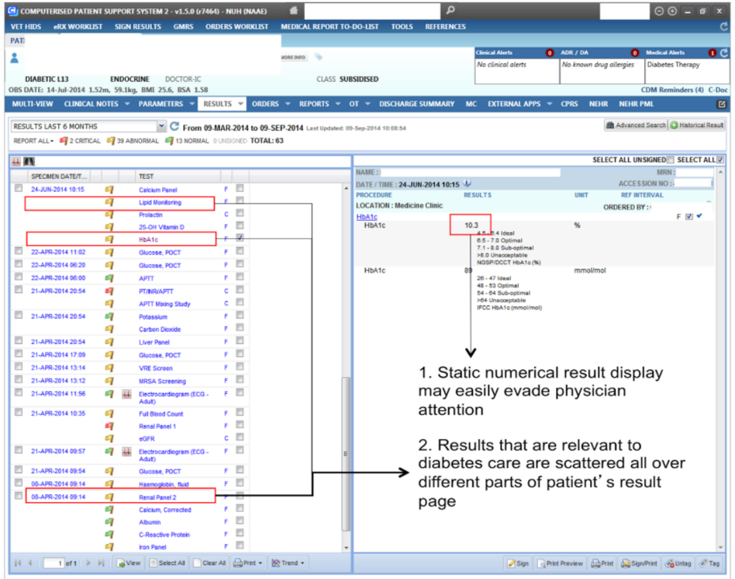
Lee, Y.-H., Bang, H., & Kim, D. J. (2016)



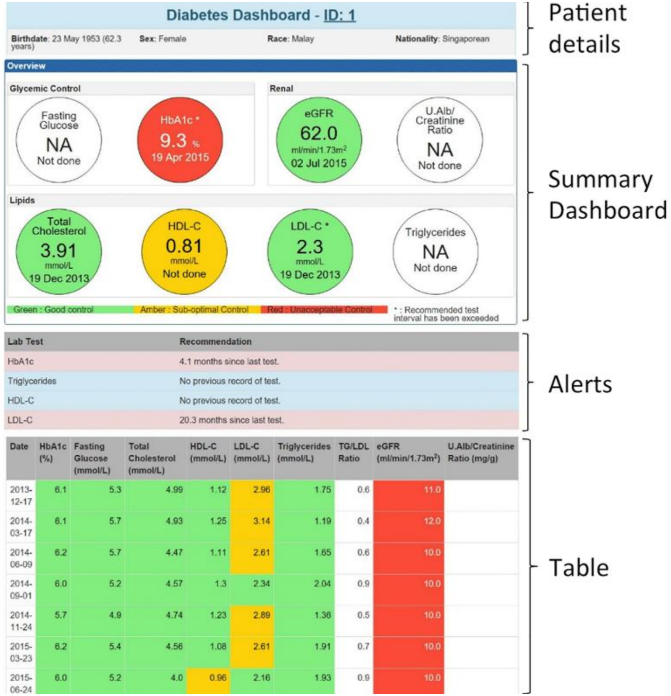
https://de.mathworks.com/help/stats/machine-learning-in-matlab.html

Clinical Decision Support Systems More Examples

- Diabetes Dashboard[1]



A screenshot of Computerized Patient Support System, Sim (2017)



Lab Test	Recommendation
HbA1c	4.1 months since last test.
Triglycerides	No previous record of test.
HDL-C	No previous record of test.
LDL-C	20.3 months since last test.

Date	HbA1c (%)	Fasting Glucose (mmol/L)	Total Cholesterol (mmol/L)	HDL-C (mmol/L)	LDL-C (mmol/L)	Triglycerides (mmol/L)	TG/LDL Ratio	eGFR (ml/min/1.73m²)	U.Alb/Creatinine Ratio (mg/g)
2013-12-17	6.1	5.3	4.99	1.12	2.96	1.75	0.6	11.0	
2014-03-17	6.1	5.7	4.93	1.25	3.14	1.19	0.4	12.0	
2014-06-09	6.2	5.7	4.47	1.11	2.61	1.65	0.6	10.0	
2014-09-01	6.0	5.2	4.57	1.3	2.34	2.04	0.9	10.0	
2014-11-24	5.7	4.9	4.74	1.23	2.89	1.36	0.5	10.0	
2015-03-23	6.2	5.4	4.56	1.08	2.61	1.91	0.7	10.0	
2015-06-24	6.0	5.2	4.0	0.96	2.16	1.93	0.9	10.0	

A screenshot of the summary dashboard, Sim (2017)

[1] Sim, L. L. W., Ban, K. H. K., Tan, T. W., Sethi, S. K., & Loh, T. P. (2017). Development of a clinical decision support system for diabetes care: A pilot study. PLoS ONE, 12(2), 1–15. http://doi.org/10.1371/journal.pone.0173021

Clinical Decision Support Systems More Examples

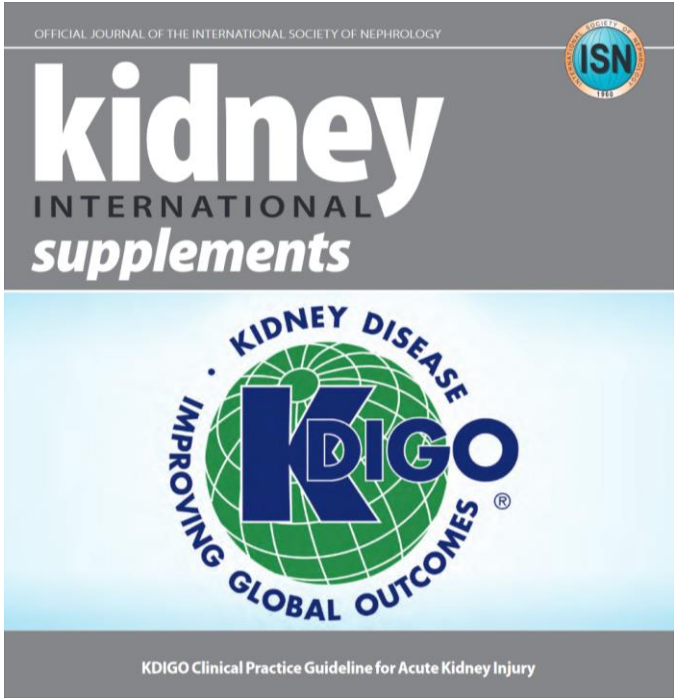
- National Kidney Foundation (US)

EGFR Calculator (NKF)

iPhone App (NKF)

https://www.kidney.org/professionals/tools

Clinical Guidelines
KDIGO – Kidney Disease Improving Global Outcomes



Kidney Disease Improving Global Outcomes, Clinical Practice Guideline for Acute Kidney Injury (2012)

Clinical Guidelines
KDIGO – Kidney Disease Improving Global Outcomes

Hasso Plattner Institut

- Recommendation level

Grade*	Implications: Patients	Implications: Clinicians	Implications: Policy
Level 1 "We recommend"	Most people in your situation would want the recommended course of action and only a small proportion would not.	Most patients should receive the recommended course of action.	The recommendation can be evaluated as a candidate for developing a policy or a performance measure.
Level 2 "We suggest"	The majority of people in your situation would want the recommended course of action, but many would not.	Different choices will be appropriate for different patients. Each patient needs help to arrive at a management decision consistent with her or his values and preferences.	The recommendation is likely to require substantial debate and involvement of stakeholders before policy can be determined.

- Quality of evidence

Grade	Quality of evidence	Meaning
A	High	We are confident that the true effect lies close to that of the estimate of the effect.
B	Moderate	The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
C	Low	The true effect may be substantially different from the estimate of the effect.
D	Very low	The estimate of effect is very uncertain, and often will be far from the truth.

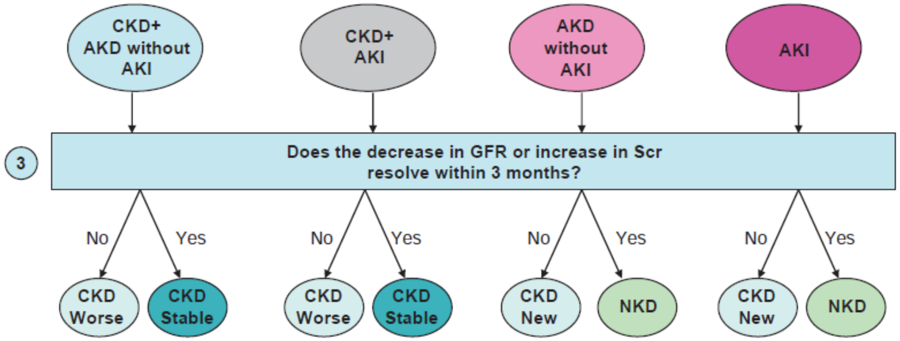
Kidney Disease Improving Global Outcomes, Clinical Practice Guideline for Acute Kidney Injury (2012)

Clinical Guidelines
KDIGO – Kidney Disease Improving Global Outcomes

- Recommendations and flowcharts

2.2.1: **We recommend that patients be stratified for risk of AKI according to their susceptibilities and exposures. (*1B*)**
2.2.2: **Manage patients according to their susceptibilities and exposures to reduce the risk of AKI (see relevant guideline sections). (*Not Graded*)**
2.2.3: **Test patients at increased risk for AKI with measurements of SCr and urine output to detect AKI. (*Not Graded*) Individualize frequency and duration of monitoring based on patient risk and clinical course. (*Not Graded*)**

KDIGO Guidelines, Risk assessment of AKI



KDIGO Guidelines, GFR/SCr algorithm: excerpt

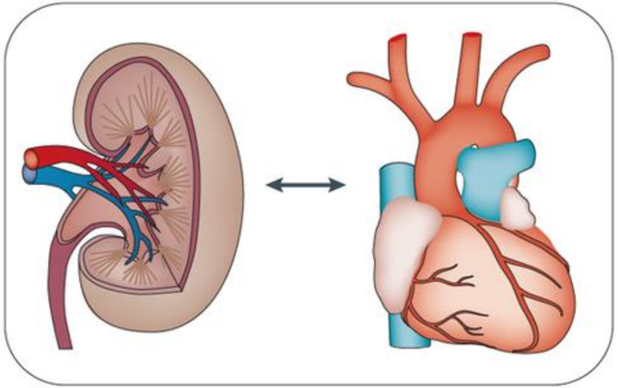
Kidney Disease Improving Global Outcomes, Clinical Practice Guideline for Acute Kidney Injury (2012)

Agenda

- Developing a Clinical Prediction Model for predicting AKI in heart surgery
- Tools for developing a Clinical Prediction Model
- Validating and evaluating models (ROC Curve)
- Use Case Nephrology exercise: what to expect?

Developing a Clinical Prediction Model
Incidence of Acute Kidney Injury in Cardiac Surgery

- Heart and kidney are deeply connected
- AKI after cardiac surgery is relatively common (3 to 30%)
- Associated with complications and mortality
- Risk factors: age, hypertension, vascular disease (i.a.)
- Cardiopulmonary bypass (CPB)
- Body fluid volume alterations
- High doses of vasopressors



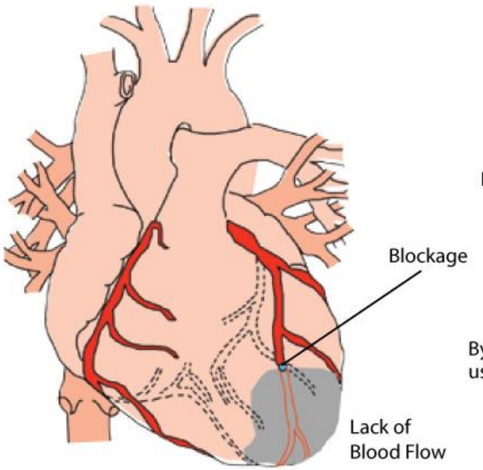
http://www.nature.com/nrneph/journal/v12/n10/abs/nrneph.2016.113.html

O'Neal, J. B., Shaw, A. D., & Billings, F. T. (2016). Acute kidney injury following cardiac surgery: current understanding and future directions. Critical Care, 20(1), 187. http://doi.org/10.1186/s13054-016-1352-z

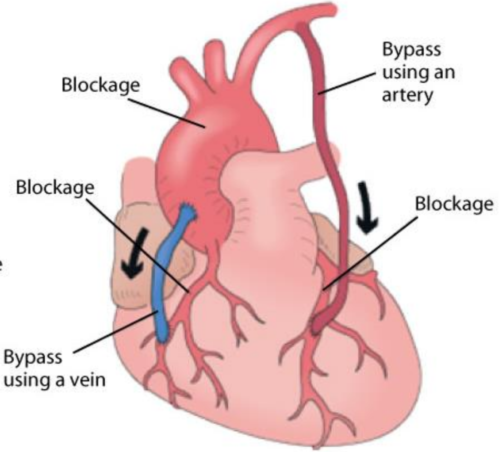
Developing a Clinical Prediction Model
Bypass and Valve Surgery

- Coronary artery bypass graft (CABG)
- ICD-9-PCS: Ch. 36.1 - Bypass anastomosis for heart revascularization

Heart attack



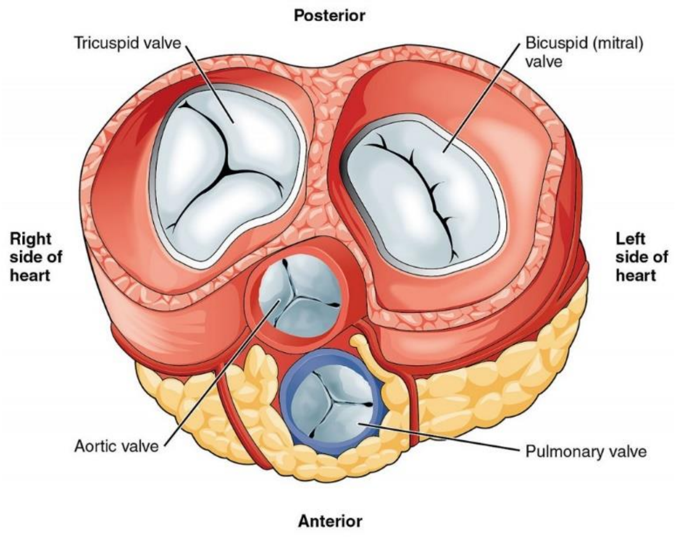
Coronary bypass



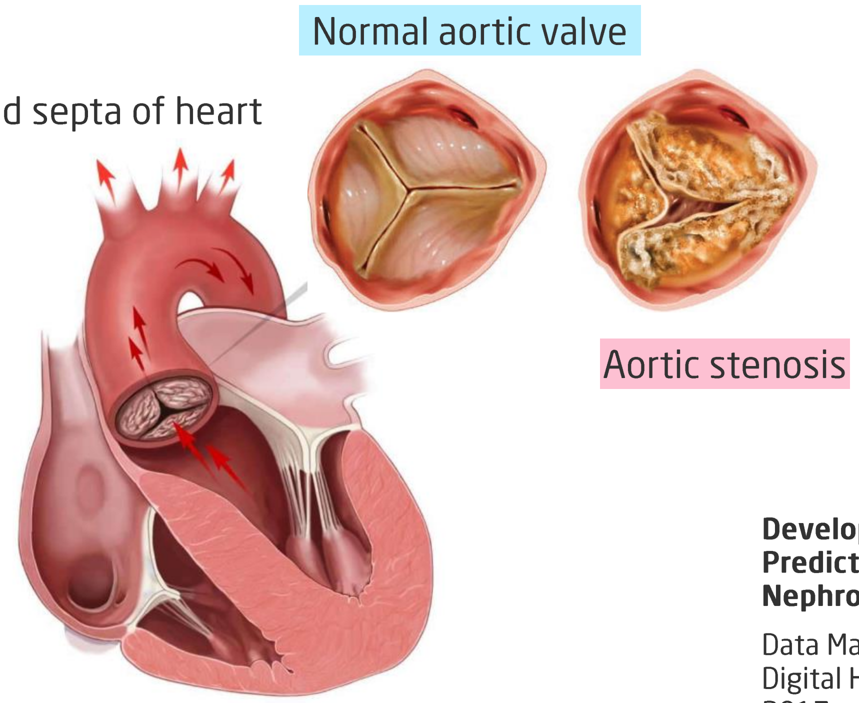
http://ctsurgery.com/heart-surgery/

Developing a Clinical Prediction Model
Bypass and Valve Surgery

- Valve surgery
- ICD-9-PCS: Ch. 35 - Operations on valves and septa of heart



http://teachmeanatomy.info/thorax/organs/heart/heart-valves/



http://www.corevalve.com/us/what-severe-aortic-stenosis/index.htm

Developing a Clinical Prediction Model
Tools for Developing CPM

R

scikit learn

WEKA
The University of Waikato

tranSMART FOUNDATION

rapidminer

Open for Innovation
KNIME®

sas®

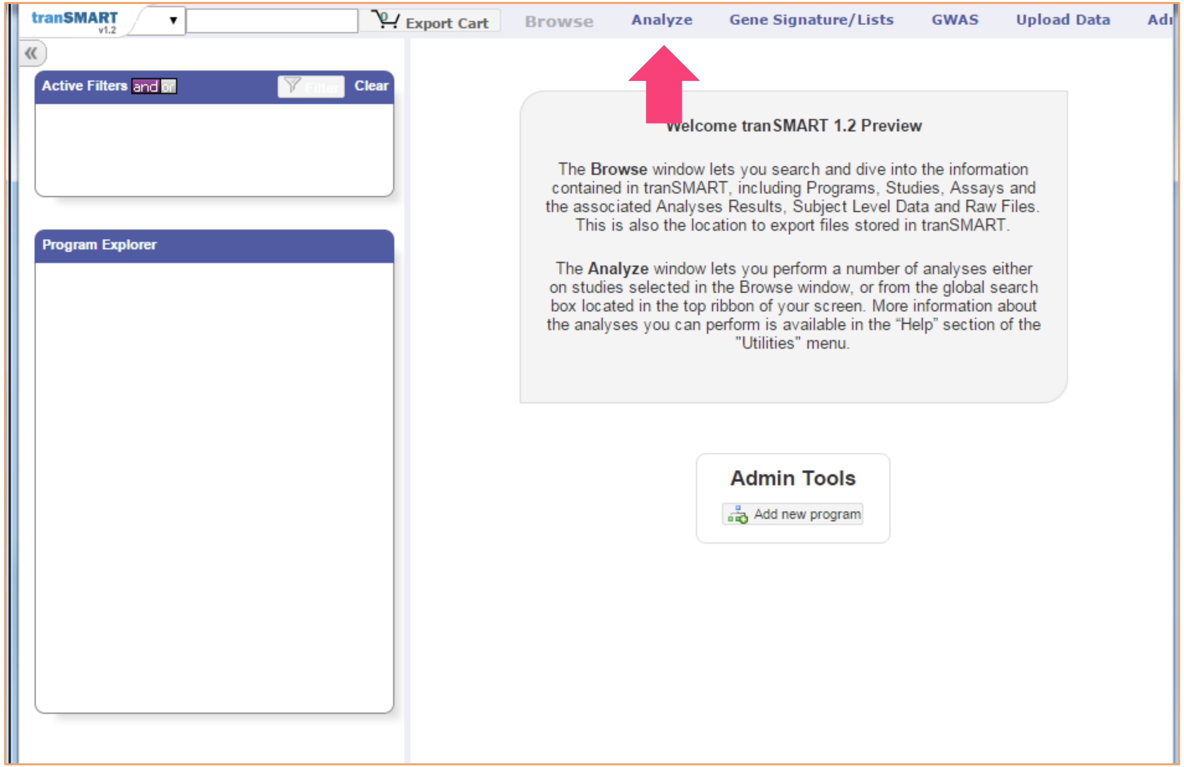
IBM SPSS

Developing a Clinical Prediction Model Tools: tranSMART

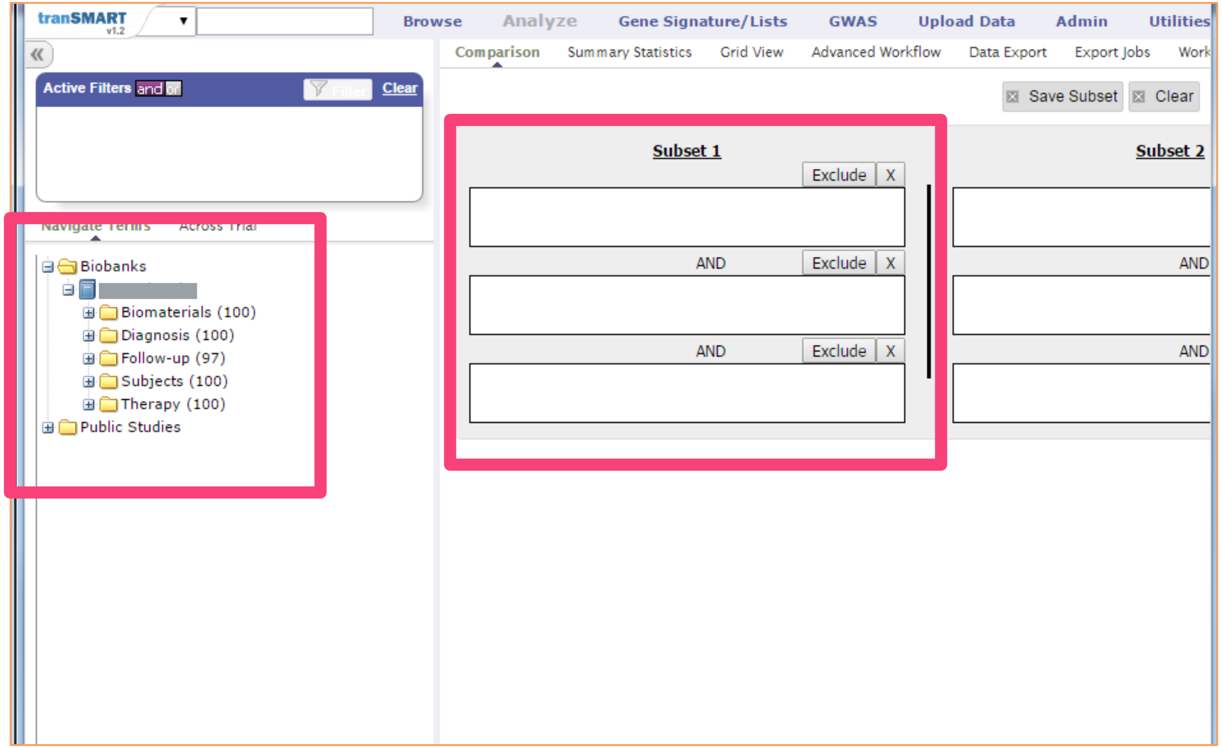
- Clinical and omics data repository
- Initial development at J&J
- Open-source project at the tranSMART Foundation
- Platform for biomedical computation
  - Clinical trials + omics
- Types of analyses (excerpt)
  - Correlation / regression
  - K-means clustering,
  - Survival analysis, aCGH, GWAS

http://transmartfoundation.org/

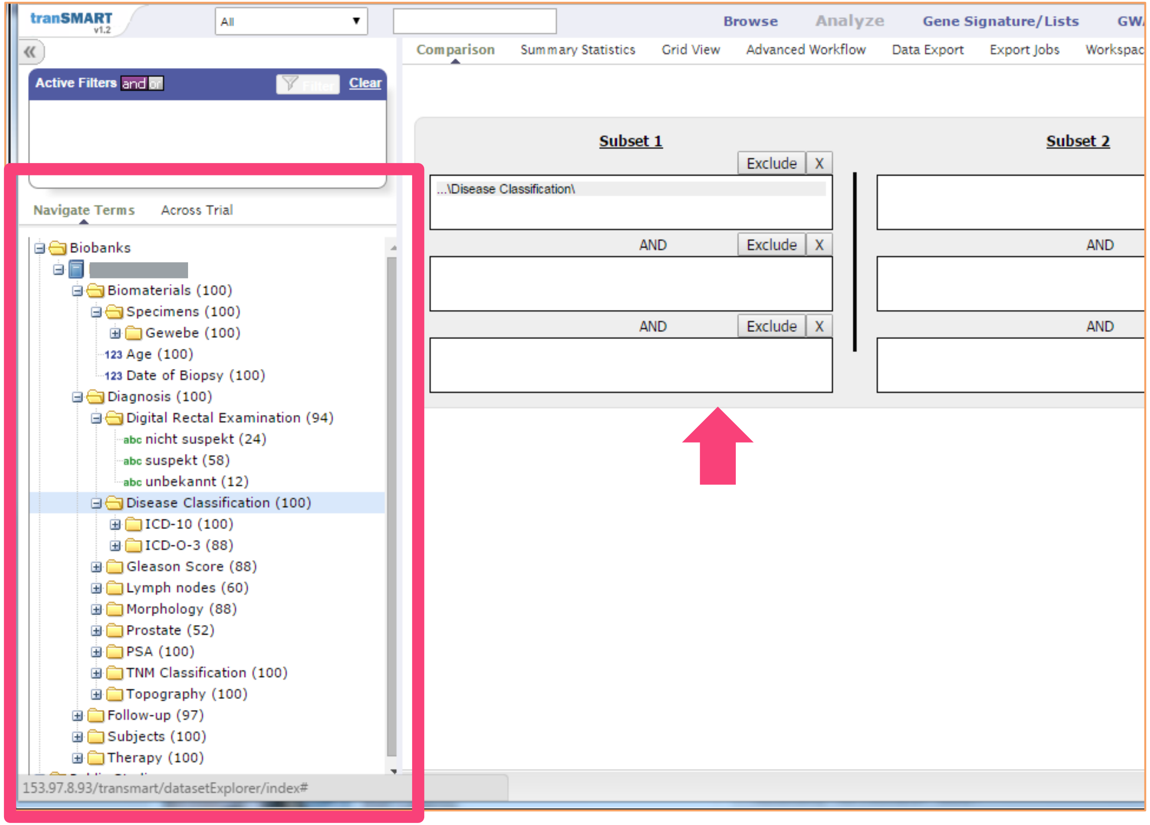
Developing a Clinical Prediction Model
Tools: tranSMART



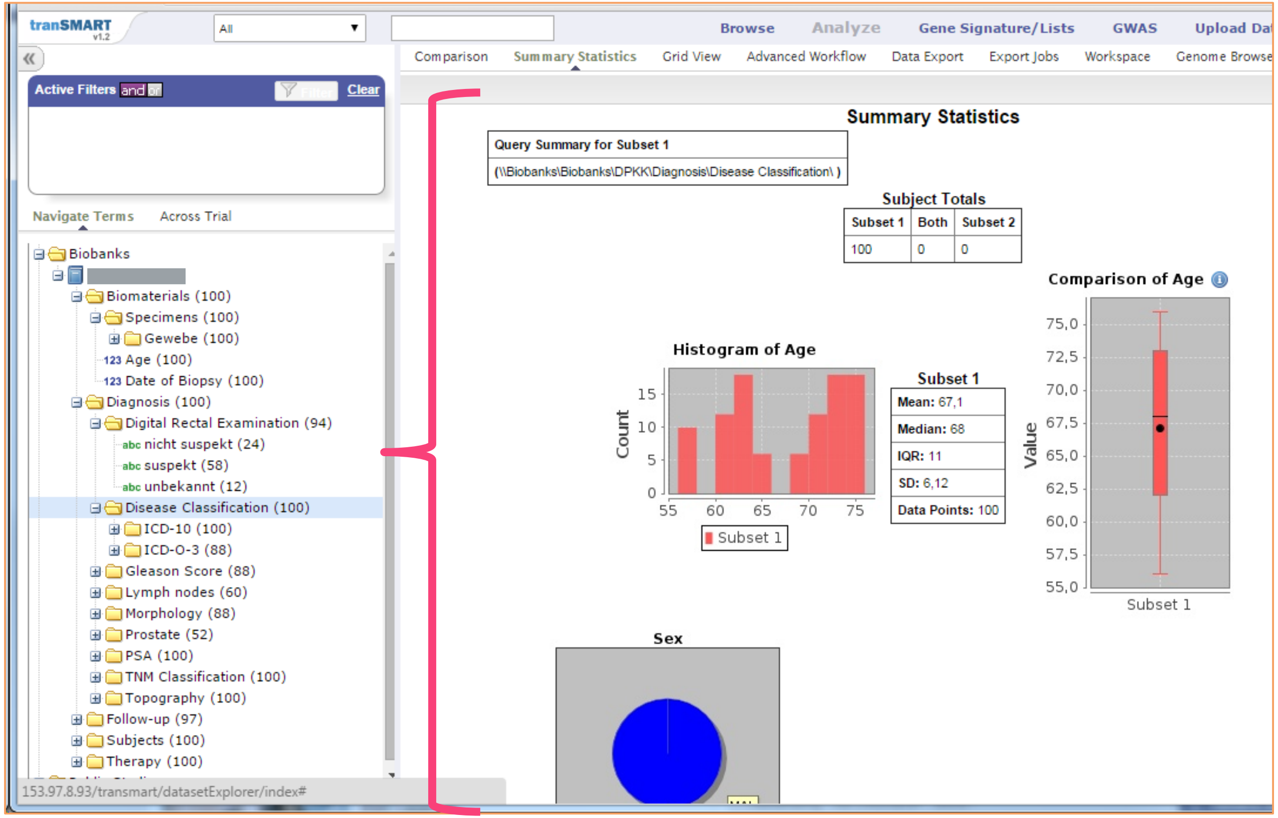
Developing a Clinical Prediction Model Tools: tranSMART



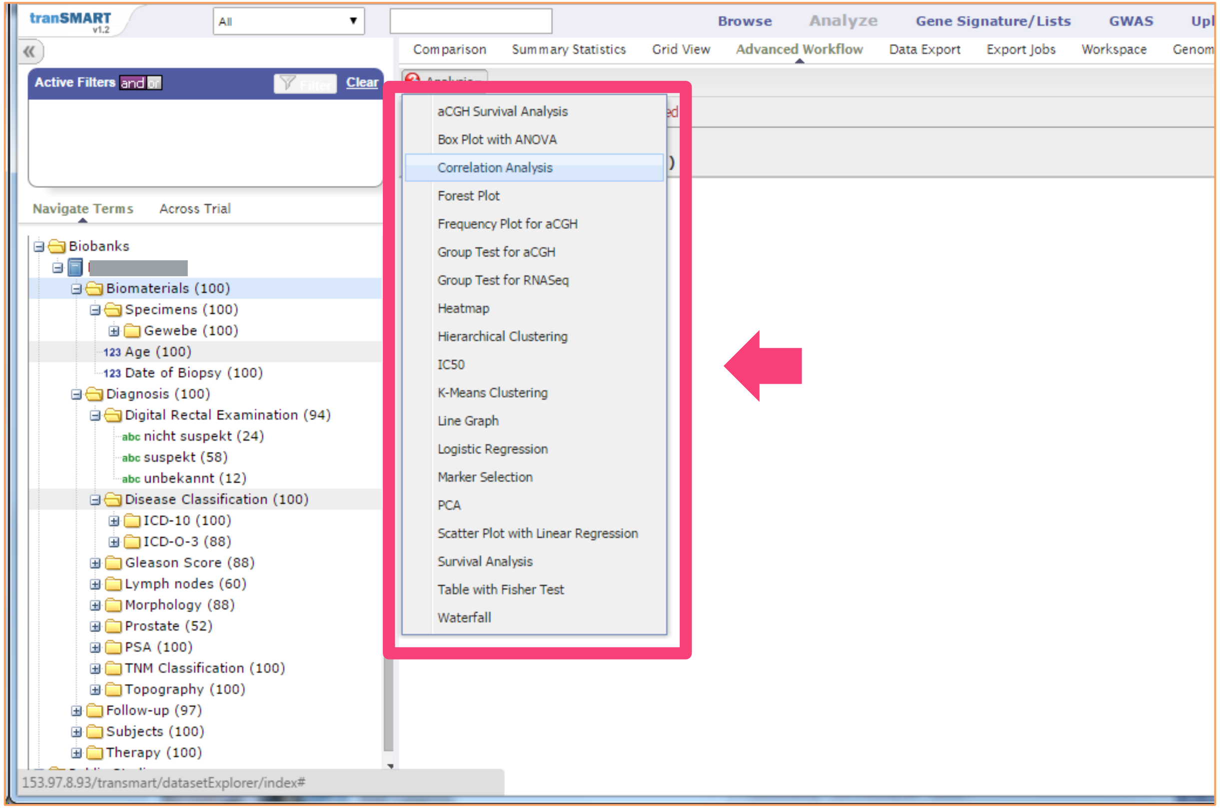
Developing a Clinical Prediction Model
Tools: tranSMART



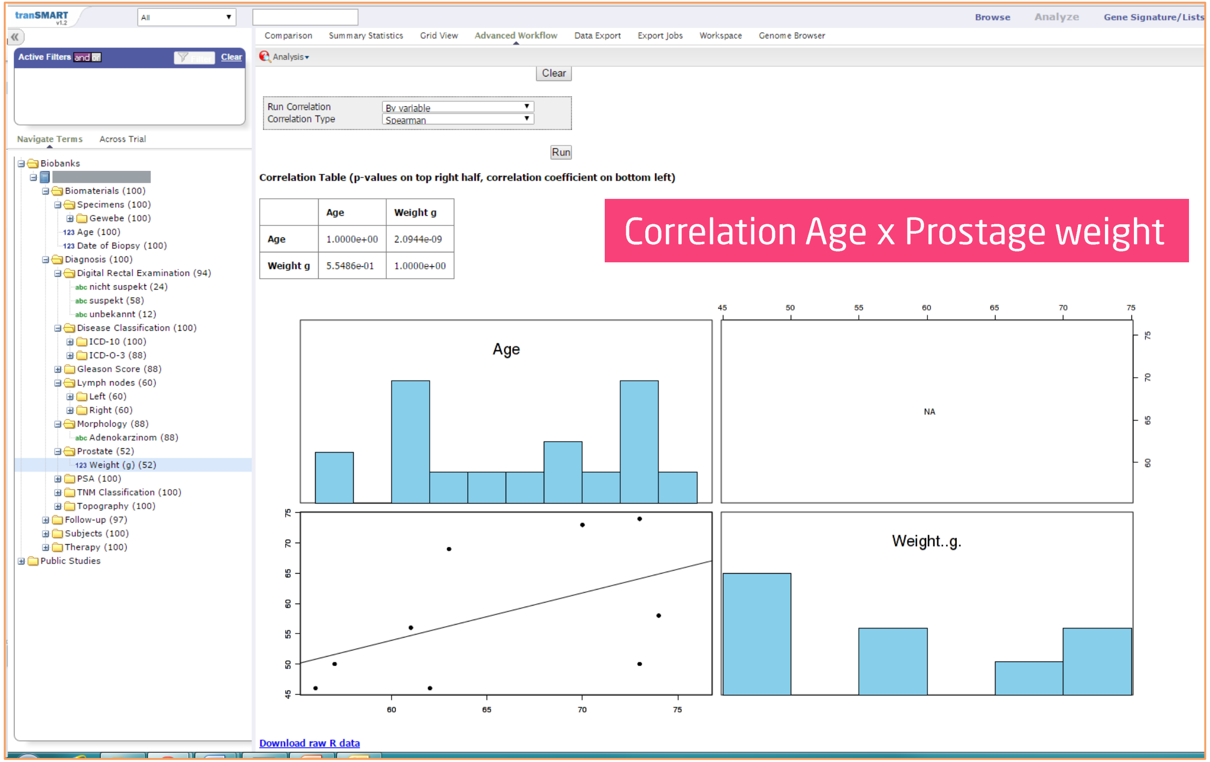
Developing a Clinical Prediction Model Tools: tranSMART



Developing a Clinical Prediction Model
Tools: tranSMART

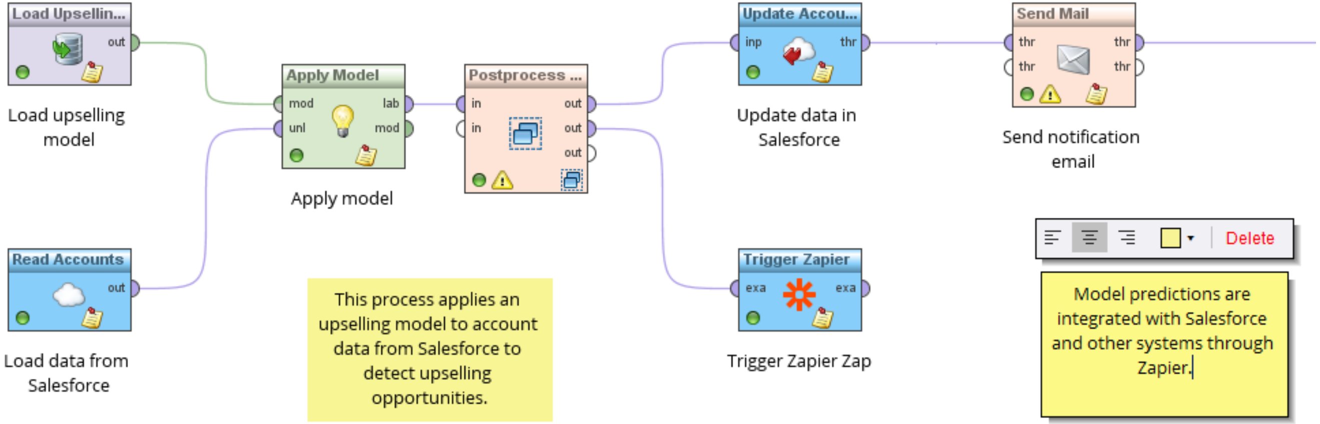


Developing a Clinical Prediction Model
Tools: tranSMART



Developing a Clinical Prediction Model
Tools: RapidMiner

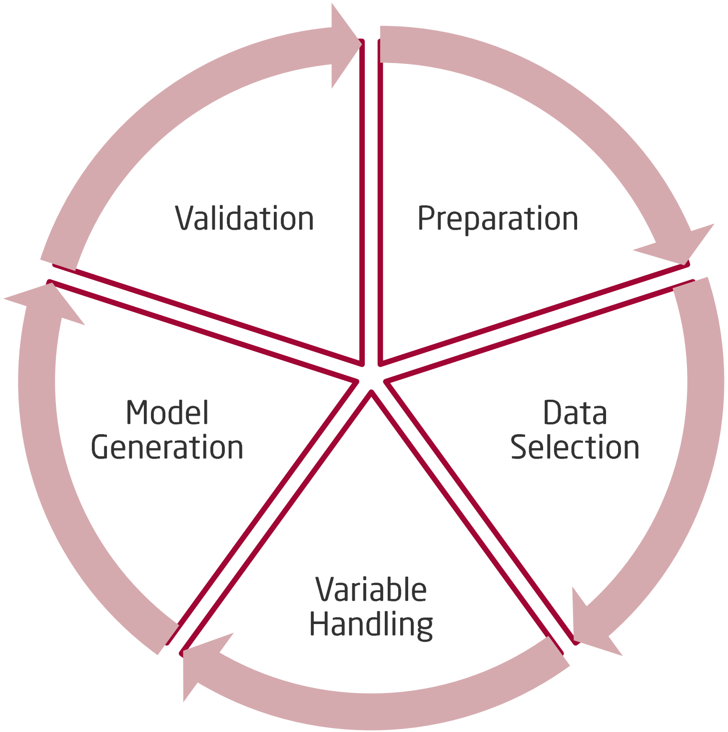
- General purpose ML-prototyping tool
- Process-based (operators)
- Data acquisition, preparation, model development and validation



https://docs.rapidminer.com/studio/releases/changes-6.4.000.html

https://rapidminer.com/us/

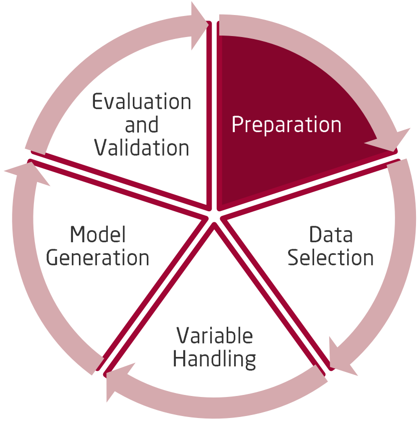
Developing a Clinical Prediction Model
General approach



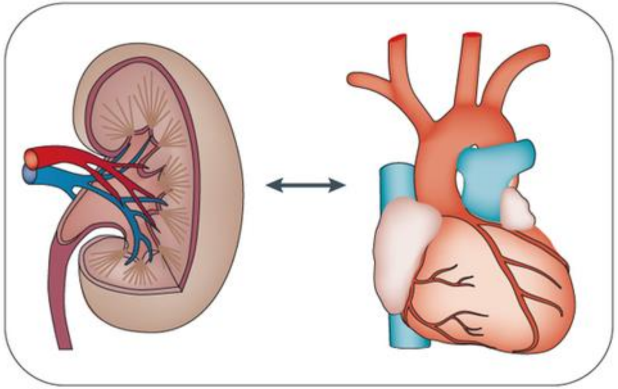
Lee, Y.-H., Bang, H., & Kim, D. J. (2016). How to Establish Clinical Prediction Models. Endocrinology and Metabolism, 31(1), 38.
http://doi.org/10.3803/EnM.2016.31.1.38

Developing a Clinical Prediction Model Preparation

- Target outcome := predict AKI in patients who had cardiac surgery
- Target patients := patients who had cardiac surgery with AKI
- Target users := cardiologists (expert users)



Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

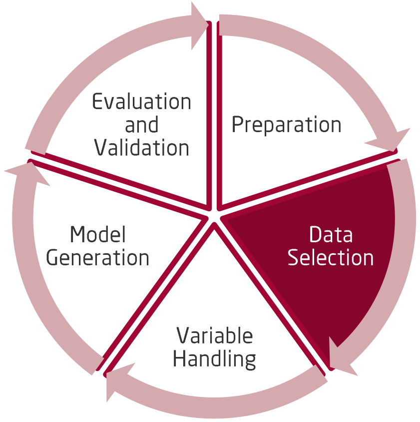


http://www.nature.com/nrneph/journal/v12/n10/abs/nrneph.2016.113.html

Developing a Clinical Prediction Model
Data Selection

- Patients who underwent *cardiac surgery*
- Patients who presented *AKI*
- Available data source := MIMICIII database[1]

Procedure / Diagnosis	Samples
Cardiac surgery	13,083
Bypass surgery	9,274
Valve surgery	3,809
Acute Kidney Injury (AKI)	1,262
Prevalence	9,65%

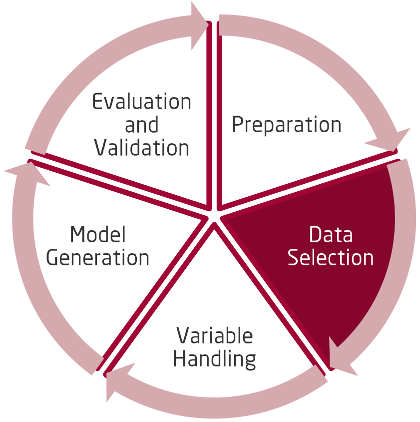


Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. http://doi.org/10.1038/sdata.2016.35

Developing a Clinical Prediction Model
Data Selection

- Data extracted
  - Demographics: gender, age, ethnicity
  - Lab values (min, max): albumin, creatinine, urea, etc.
  - Comorbidities: diabetes, hypertension, obesity, etc.
  - ICU scores: SOFA, SAPS, OASIS
- In total: 78 attributes



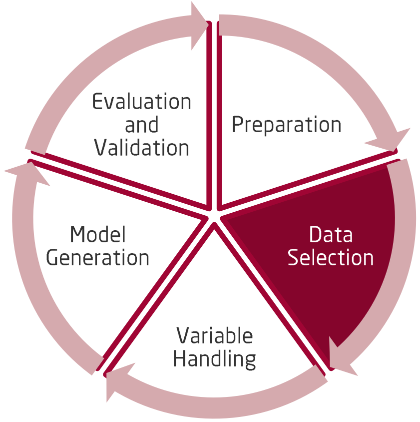
Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. http://doi.org/10.1038/sdata.2016.35

Developing a Clinical Prediction Model
Data Selection: SQL is your friend!

```sql
select  pat.subject_id,    adm.hadm_id,    pat.gender,    adm.ethnicity,  round(days_between(pat.dob
, adm.admittime)/365,0) as age, elix_score.elixhauser_vanwalraven,       aids,   alcohol_abuse,
    blood_loss_anemia,  cardiac_arrhythmias,    chronic_pulmonary,  coagulopathy,
    congestive_heart_failure,  deficiency_anemias,    depression,    diabetes_complicated,
    diabetes_uncomplicated,    drug_abuse,    fluid_electrolyte,  hypertension,   hypothyroidism,
        liver_disease,  lymphoma,   metastatic_cancer,  obesity,    other_neurological,     paralysis,
     peptic_ulcer,   peripheral_vascular,    psychoses,  pulmonary_circulation,  renal_failure,
    rheumatoid_arthritis,   solid_tumor,    valvular_disease,   weight_loss, sofa.sofa as sofa_total,
      sofa.renal as sofa_renal,    saps.saps as saps_total,    oasis.oasis as oasis_total,
    albumin_max,    albumin_min,    aniongap_max,   aniongap_min,   bands_max,  bands_min,
    bicarbonate_max,    bicarbonate_min,    bilirubin_max,  bilirubin_min,  bun_max,    bun_min,
    chloride_max,   chloride_min,   creatinine_max,     creatinine_min,     glucose_max,
    glucose_min,    hematocrit_max,     hematocrit_min,     hemoglobin_max,     hemoglobin_min,
    inr_max,    inr_min,    lactate_max,    lactate_min,    platelet_max,   platelet_min,
    potassium_max,  potassium_min,  pt_max,     pt_min,     ptt_max,    ptt_min,    sodium_max,
    sodium_min,     wbc_max,    wbc_min,    hsa.aki,    case        when (hsa.icd_proc between 3610 and
     3619) then 'Bypass surgery'         when (hsa.icd_proc between 3500 and 3599) then 'Valve surgery'
         end         from mimiciii.admissions adm    inner join mimiciii.patients pat        on adm.
    subject_id = pat.subject_id          inner join mimiciii.m_elixhauser_quan elix      on elix.
    hadm_id = adm.hadm_id            inner join mimiciii.m_elixhauser_quan_score elix_score      on
    elix_score.hadm_id = adm.hadm_id            inner join                  (select * from (select
    sofa.*,         min(icustay_id) over(partition by subject_id, hadm_id) as min_icustay_id
         from mimiciii.m_sofa sofa          ) where icustay_id = min_icustay_id ) sofa on sofa.hadm_id
    = adm.hadm_id        inner join             (select * from (select saps.*,          min(
    icustay_id) over(partition by subject_id, hadm_id) as min_icustay_id            from mimiciii.
    m_saps saps         ) where icustay_id = min_icustay_id ) saps on saps.hadm_id = adm.hadm_id
            inner join          (select * from (select oasis.*,             min(icustay_id)
    over(partition by subject_id, hadm_id) as min_icustay_id            from mimiciii.m_oasis oasis
          ) where icustay_id = min_icustay_id ) oasis on oasis.hadm_id = adm.hadm_id      inner join
              (select * from (select labs.*,          min(icustay_id) over(partition by
    subject_id, hadm_id) as min_icustay_id          from mimiciii.m_labsfirstday labs           )
    where icustay_id = min_icustay_id ) labs on labs.hadm_id = adm.hadm_id      inner join
    mimiciii.heart_surgery_aki hsa      on hsa.hadm_id = adm.hadm_id  order by pat.subject_id, adm.
    hadm_id
```

AKI example: Data extraction SQL

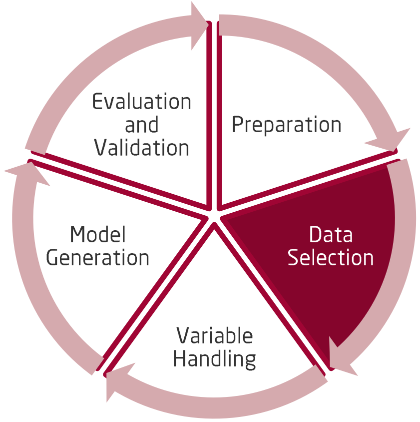


Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

Developing Clinical Prediction Models for Nephrology

Developing a Clinical Prediction Model
Data Selection

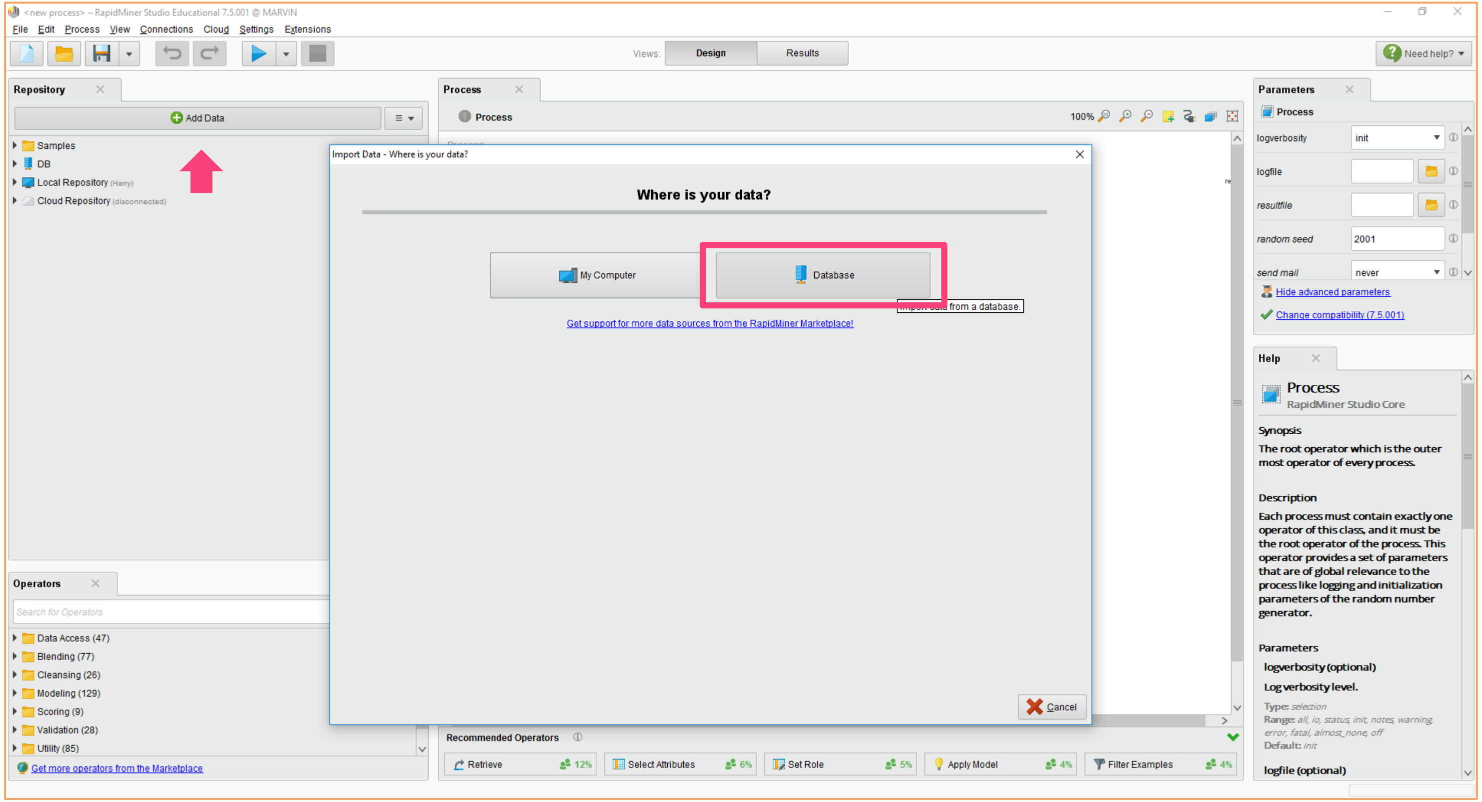
- Data extracted
  - Demographics: gender, age, ethnicity
  - Lab values (min, max): albumin, creatinine, urea, etc.
  - Comorbidities: diabetes, hypertension, obesity, etc.
  - ICU scores: SOFA, SAPS, OASIS
- In total: 78 attributes



Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

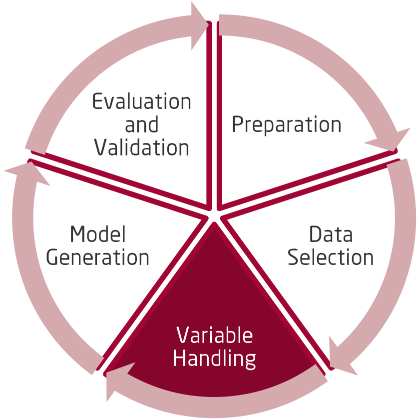
Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. http://doi.org/10.1038/sdata.2016.35

Developing a Clinical Prediction Model
Data Selection: Loading Data



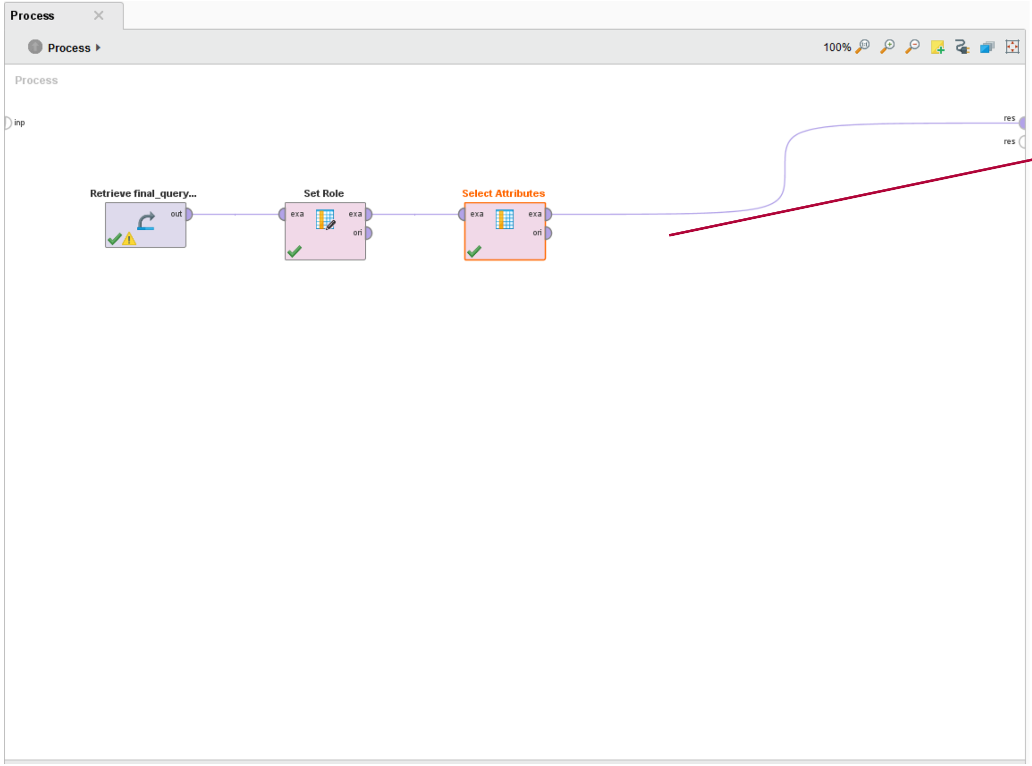
Developing a Clinical Prediction Model
Variable Handling

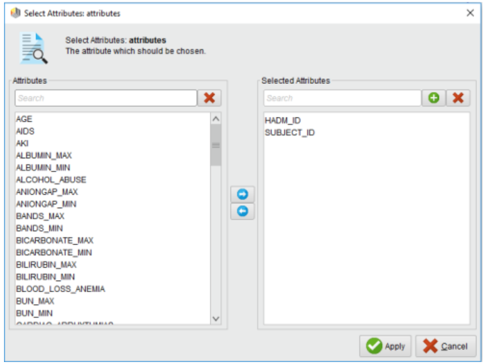
- Watch out for non-relevant predictors (e.g. IDs, etc.)
- Some variables might not be related at all with the output
- Inspect missing values: consider removing variables
- Are any data transformations needed?
  - Highly dependent on the underlying algorithm
  - Normalization and log scale
  - Nominal to numerical



Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

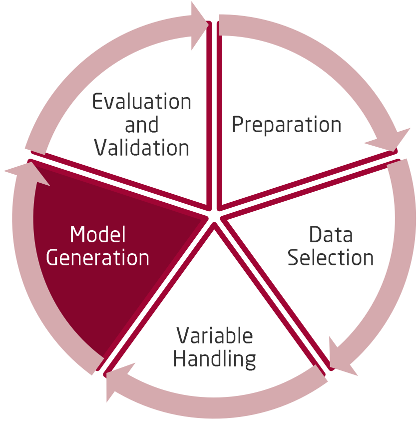
Developing a Clinical Prediction Model
Data Selection and Variable Handling





Developing a Clinical Prediction Model
Model Generation

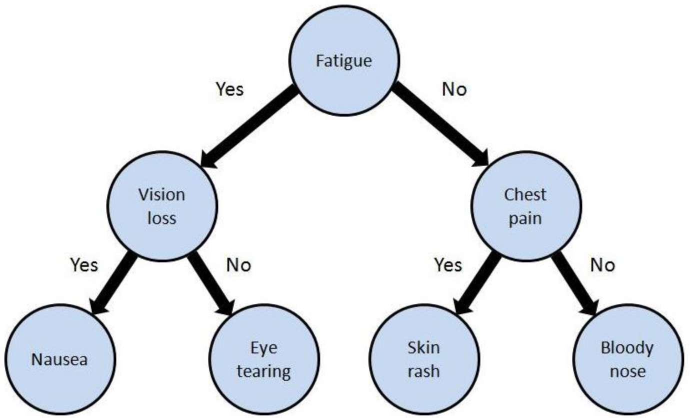
- Start with simple models (Occam‘s razor)
- Try out some algorithms and compare performance
- One of the most used in medicine: decision trees
- Consider performing feature selection
- Algorithm parameter tuning



Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

Developing a Clinical Prediction Model
Model Generation: Decision Trees

- Decision rules inferred from data
- Advantages:
  - Interpretability („white box“)
  - Can often be combined with other algorithms
  - Requires little data preparation
- Disadvantages
  - As dimensions increase, so does complexity
  - May lack generalization, prone to overfitting
  - Creates bias if classes are unbalanced

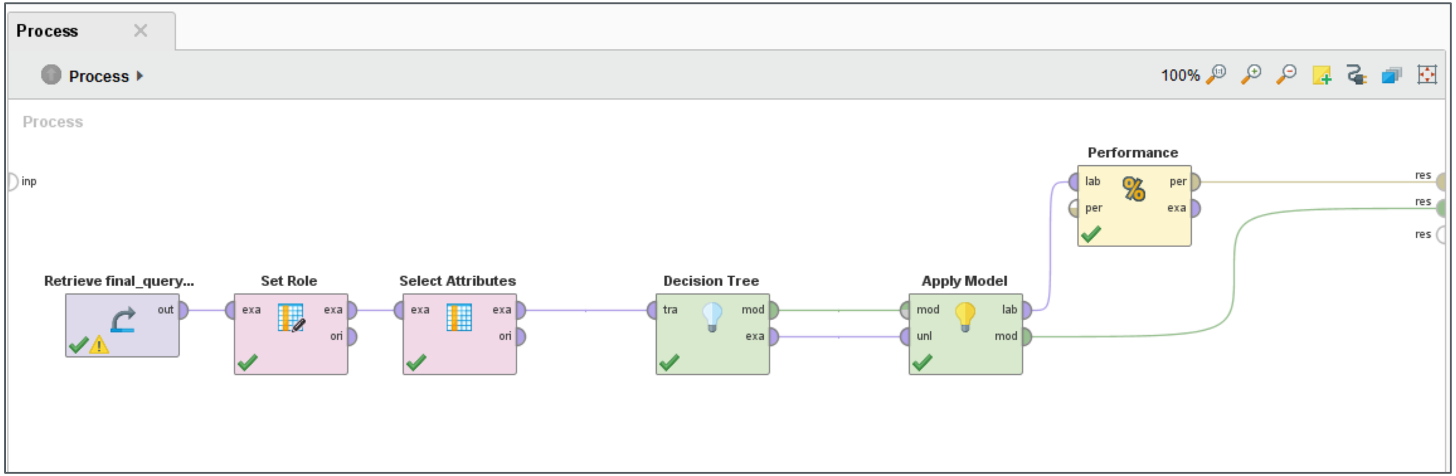


http://web.eecs.umich.edu/~cscott/research/decision_tree.jpg

http://scikit-learn.org/stable/modules/tree.html

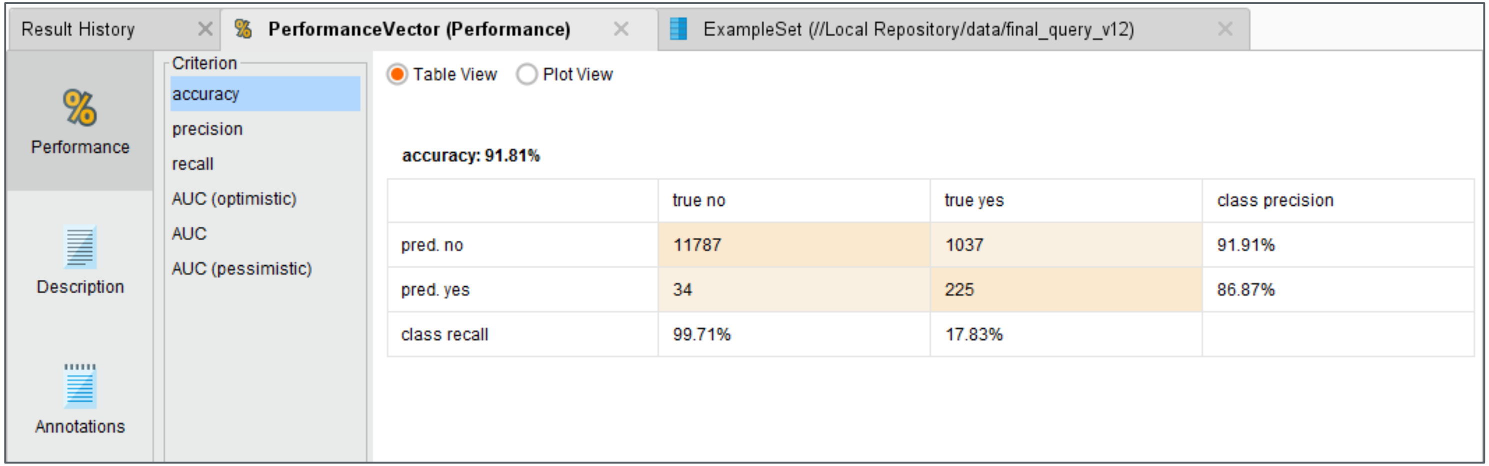
Developing a Clinical Prediction Model
Model Generation: RapidMiner

- Decision tree with no validation



Developing a Clinical Prediction Model
Model Generation: RapidMiner

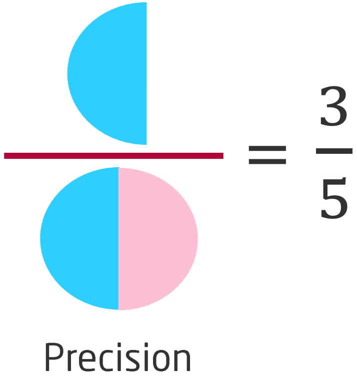
- Model performance: precision and recall?

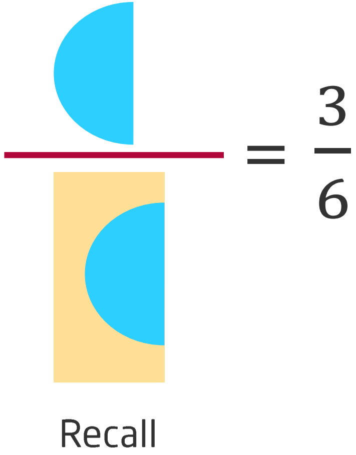


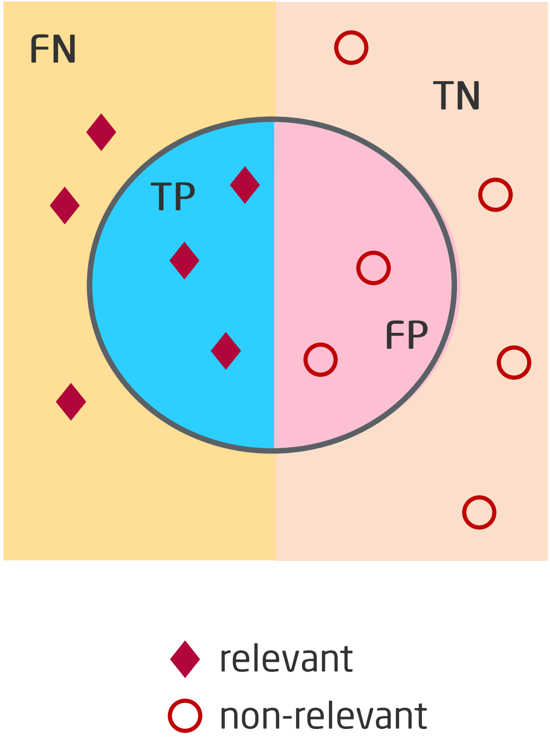
	true no	true yes	class precision
pred. no	11787	1037	91.91%
pred. yes	34	225	86.87%
class recall	99.71%	17.83%	

Clinical Prediction Models
Precision and Recall

- Both are relevance measures
- Precision := % retrieved instances that are relevant
- Recall := % relevant documents in the result set





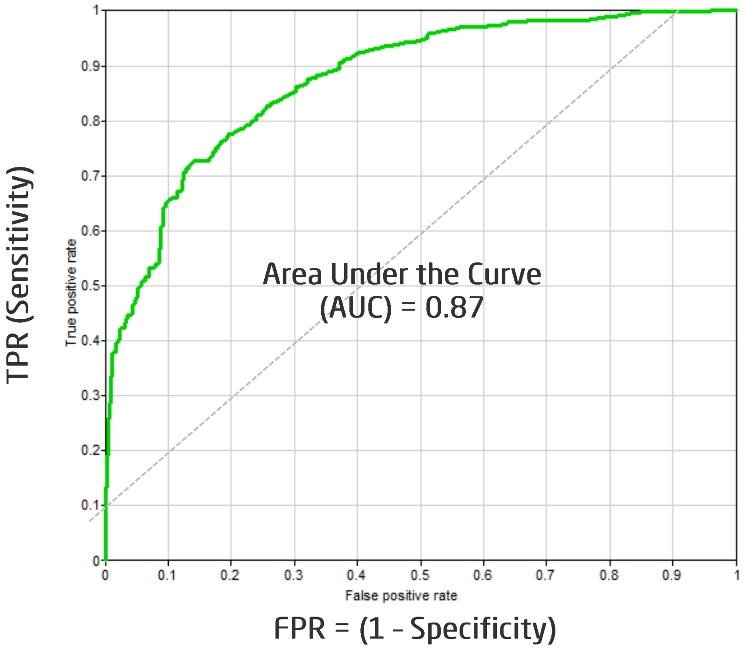


Clinical Prediction Models
Confusion Matrix

- Precision := $\dfrac{TP}{TP + FP}$
- Recall := $\dfrac{TP}{TP + FN}$
- F-measure := $2 \cdot \dfrac{Precision \cdot Recall}{Precision + Recall}$
- Sensitivity := True Positive Rate := Recall
- Specificity := True Negative Rate

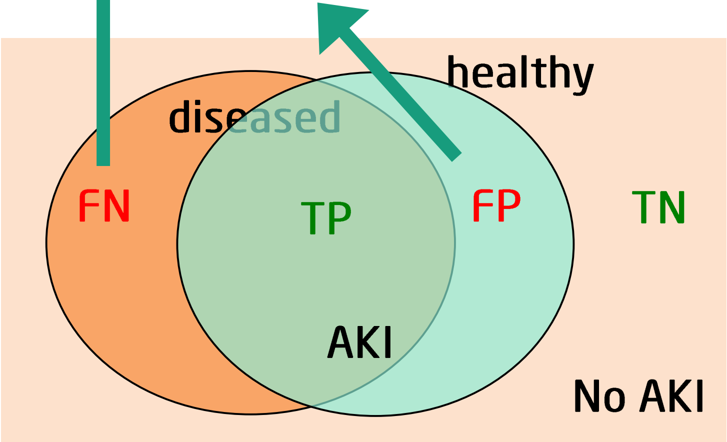
		Predicted	
		Positive	**Negative**
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

AKI prediction (classifier accuracy)
Receiver Operating Characteristic (ROC) curve



recall (sensitivity) = TP / (TP + FN)

precision = TP / (TP + FP)

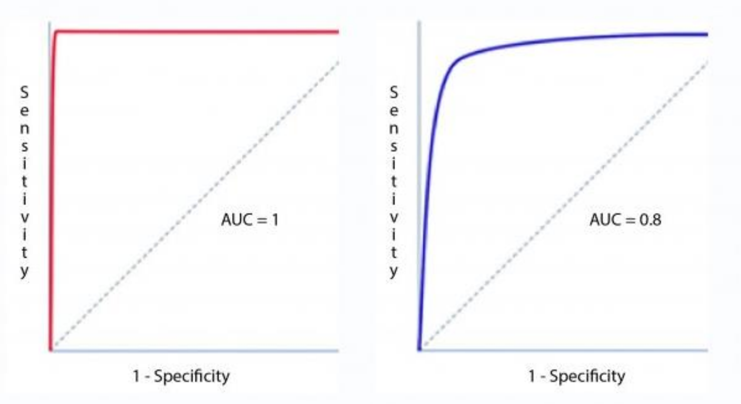


specificity = TN / (TN + FP)

Clinical Prediction Models
Receiver Operating Characteristic (ROC) Curve

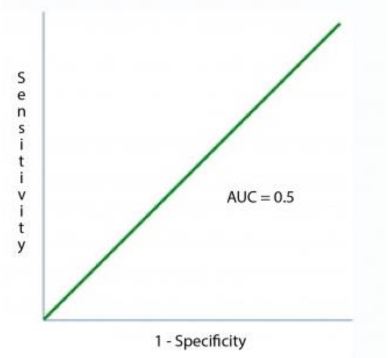
Hasso Plattner Institut

- Performance of a binary classifier
- Plot showing TPR and FPR
- Varying *classification thresholds*
- Allows comparison between classifiers
- The higher the AUC (Area Under the Curve) the better



$$TPR = \frac{TP}{\underbrace{TP + FN}_{\text{all positive instances}}}$$

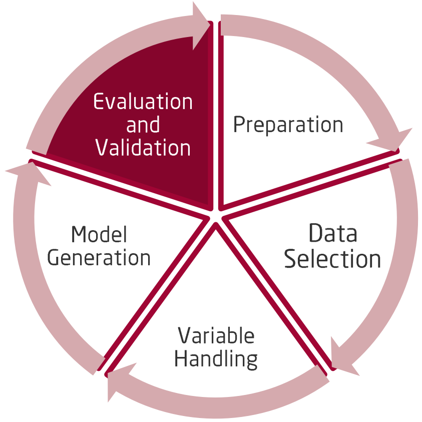
$$FPR = \frac{FP}{\underbrace{FP + TN}_{\text{all negative instances}}}$$



http://www.cienciasinseso.com/en/tag/roc-curve/

Developing a Clinical Prediction Model
Evaluation and Validation

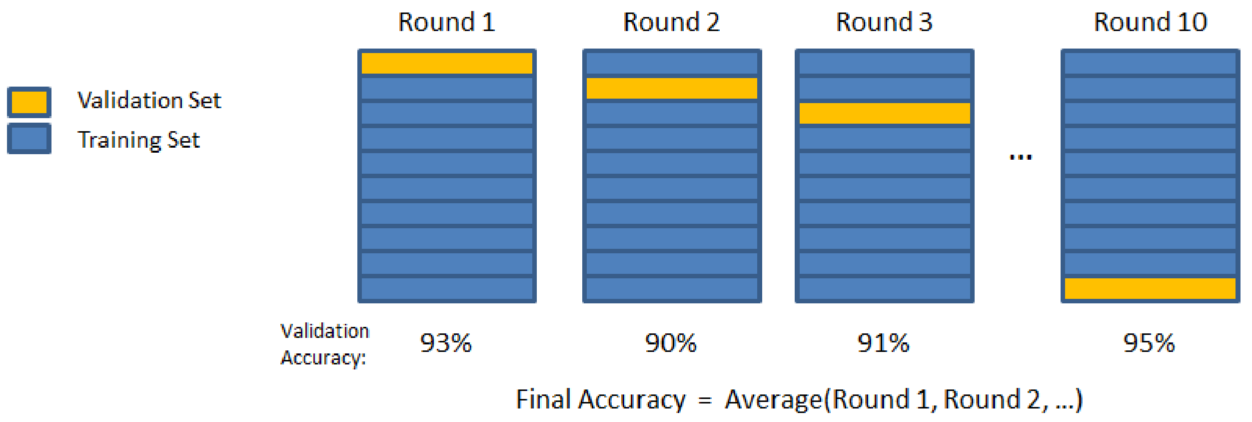
- Apply cross-validation to assess generalizability
- Try out different algorithms
- Assess the impact of missing values
  - Complete case analysis
  - Single or multiple imputation
- Assess whether class data is balanced
- Perform feature selection



Lee, Y.-H., Bang, H., & Kim, D. J. (2016)

Developing a Clinical Prediction Model
Cross-validation (*k*-fold)

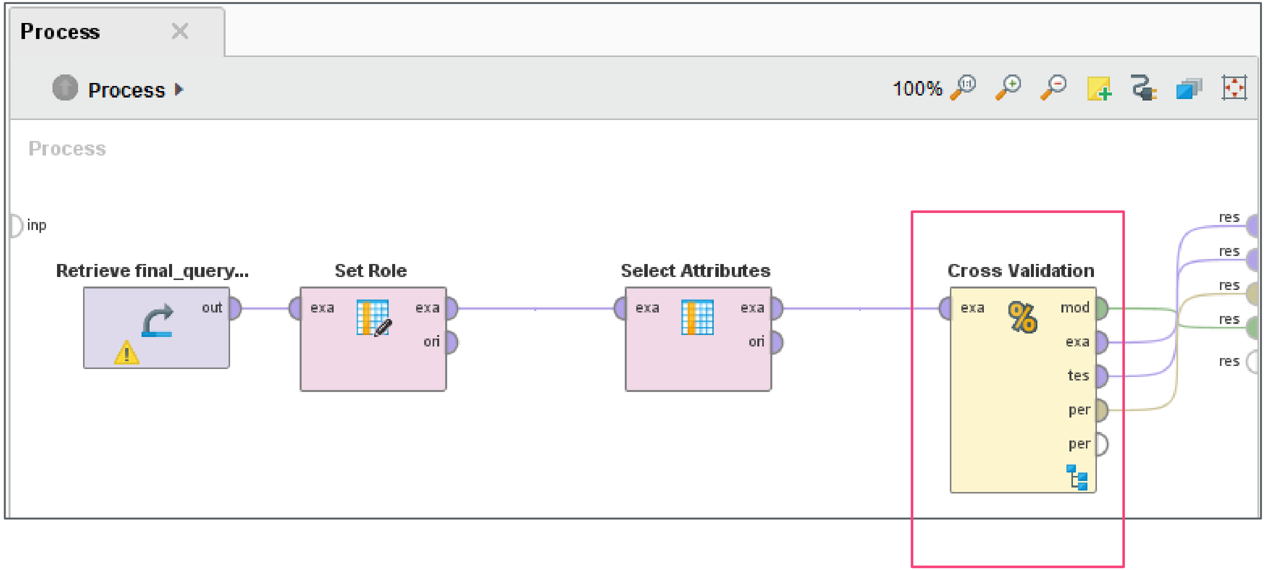
- How general is the model?
- Important to avoid *overfitting*
- Useful for internal validation of a model
- Training and validation sets
- Random split into *k* subsamples



https://chrisjmccormick.files.wordpress.com/

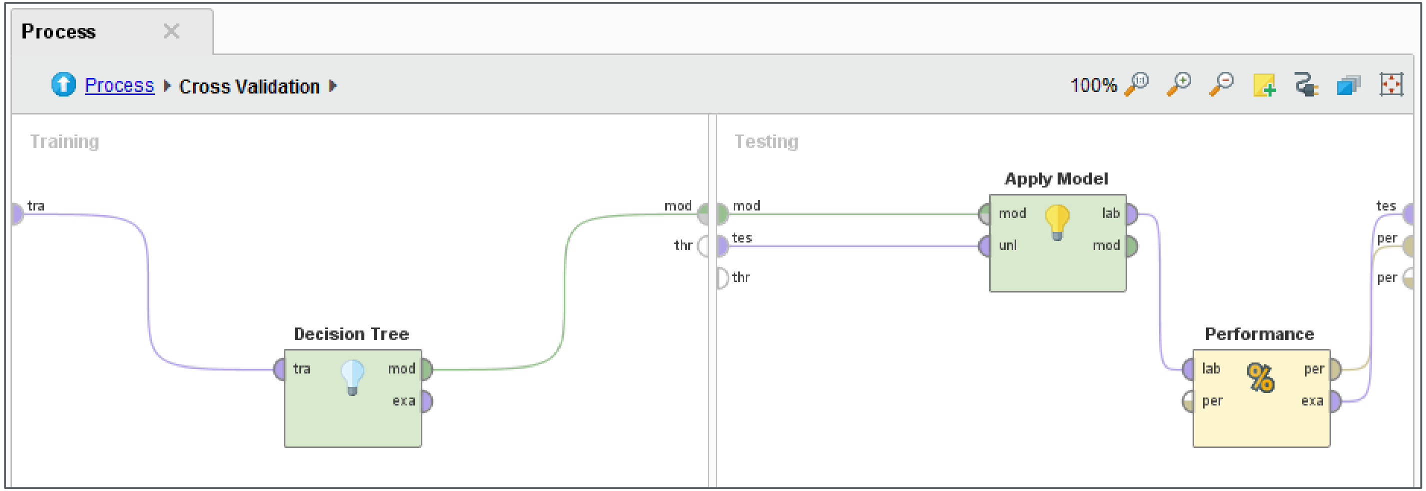
Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Cross validation operator



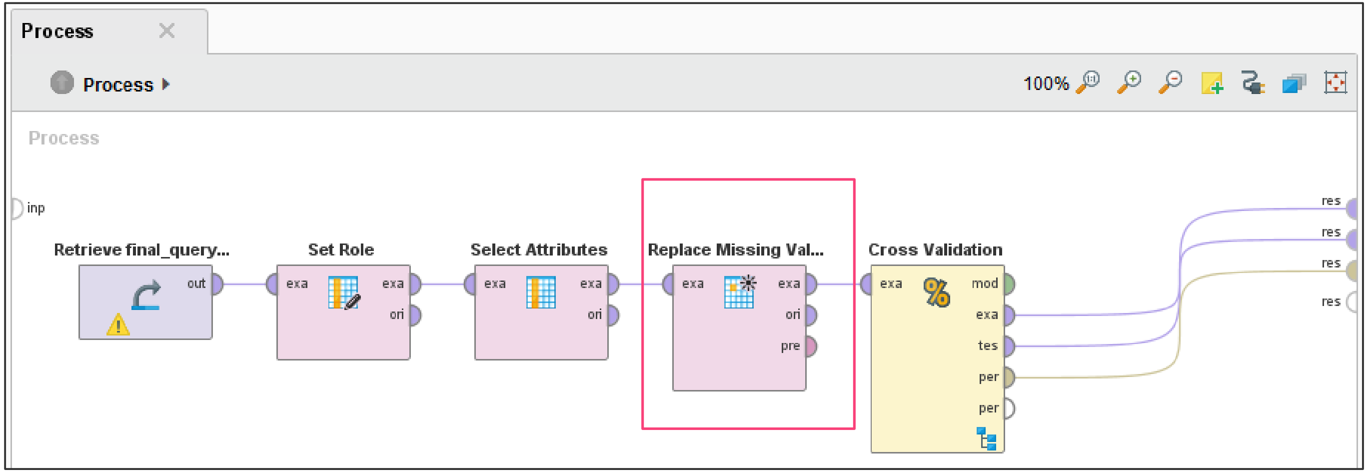
Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Cross-validation operator



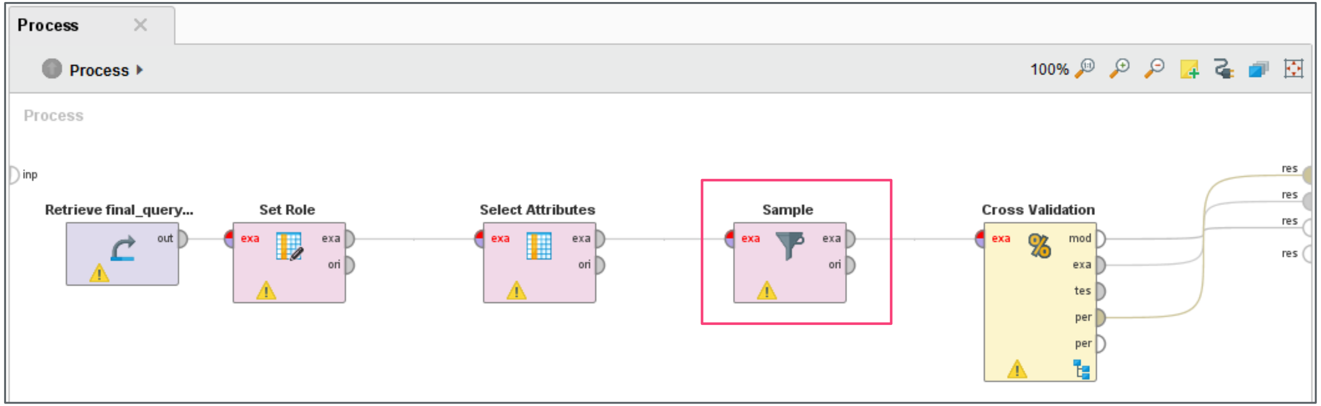
Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Replace missing values operator



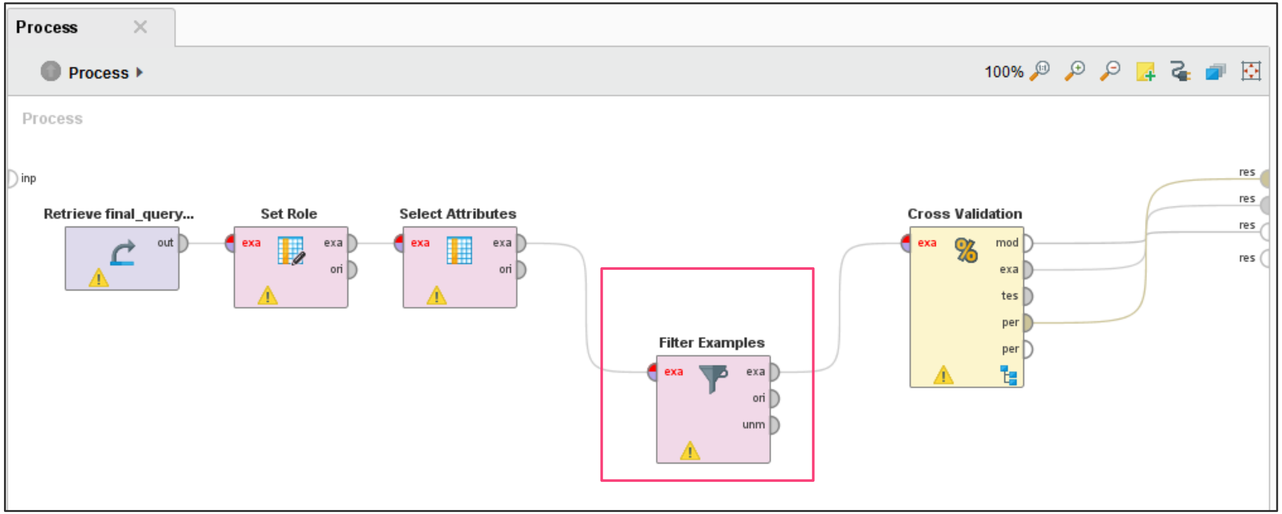
Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Sample operator: can be used for balancing datasets



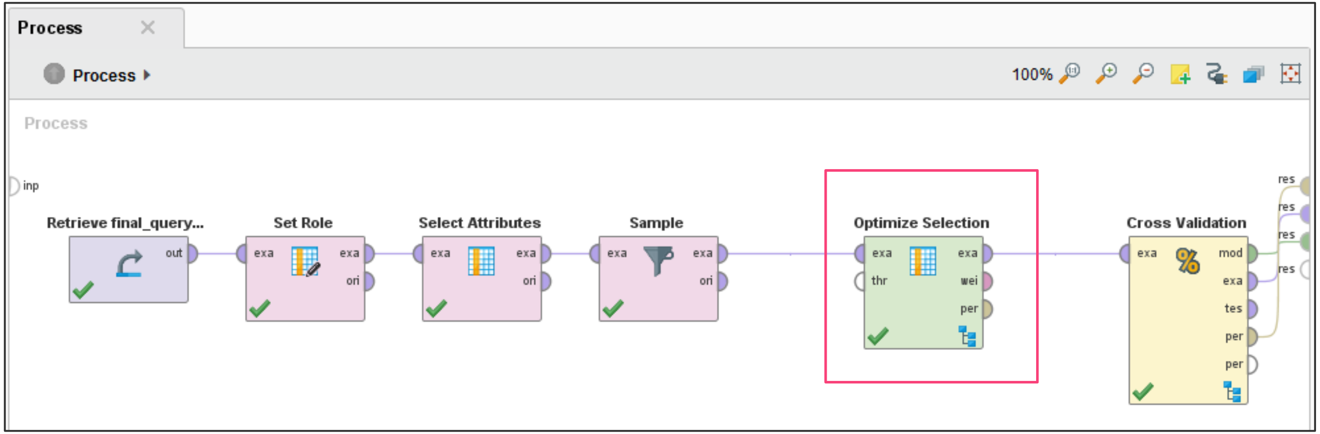
Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Filter examples operator: remove all missing data

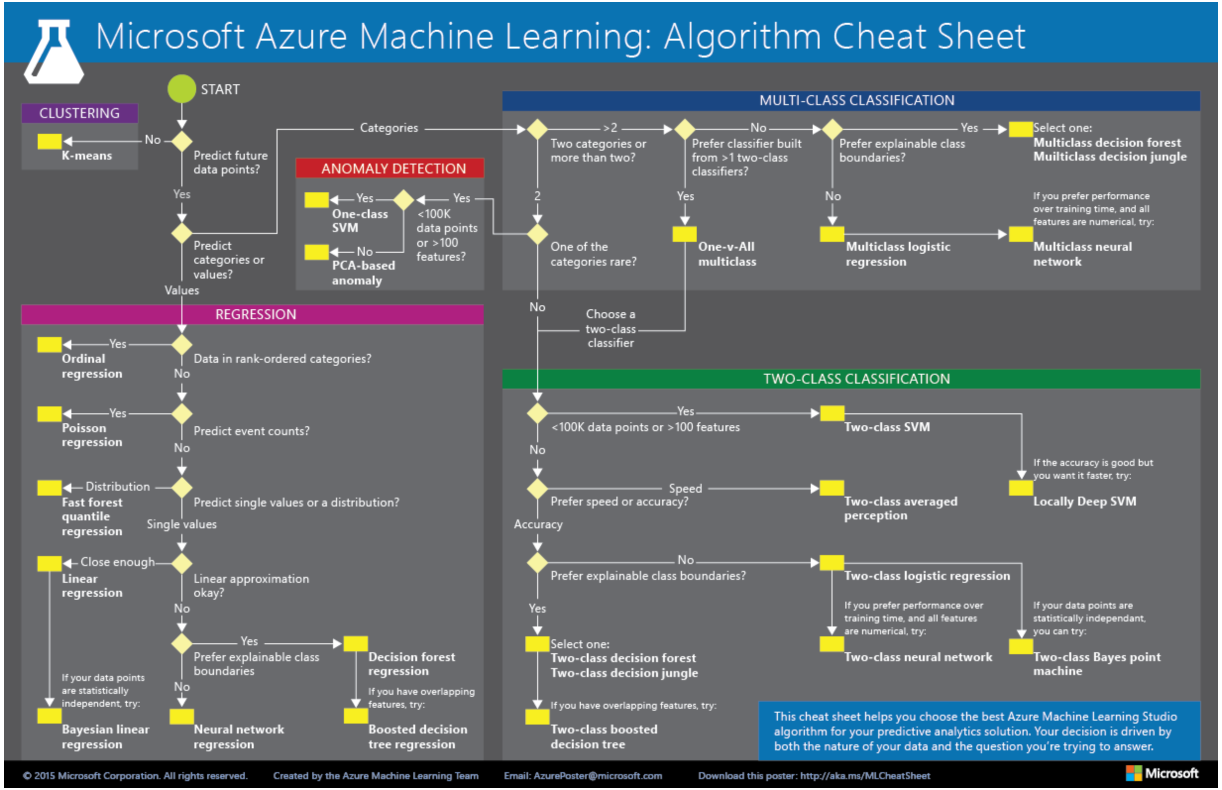


Developing a Clinical Prediction Model Evaluation and Validation: RapidMiner

- Optimize selection operator



Developing a Clinical Prediction Model
How to choose an algorithm?



http://oliviaklose.azurewebsites.net/machine-learning-11-algorithms-explained/

Exercise II
Developing a Clinical Predicion Model for Nephrology

- What you will need
  - Download and install RapidMiner
  - Obtain an academic license
  - Install JDBC HANA driver for RapidMiner
  - Connect to HANA and perform experiments
- Proceed according to the step-by-step process
  - Develop the model
  - Answer the questions on openHPI

What to Take Home?

- Developing a clinical prediction model step-by-step
- Impact of missing data on performance of the models
- Impact of class balancing on model performance
- Important concepts: precision, recall, ROC curve
- Evaluating and validating models

What's Coming Next?

- Use Case Oncology