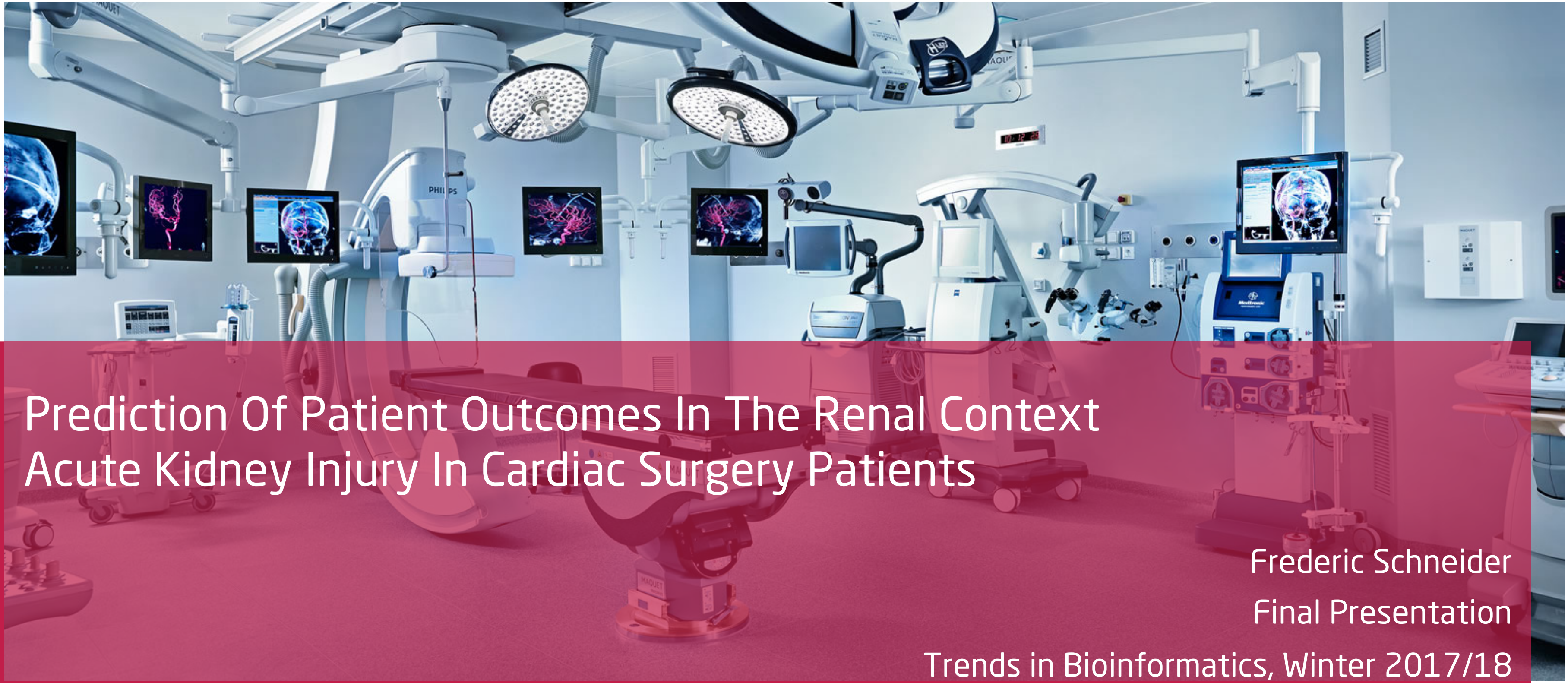


Prediction Of Patient Outcomes In The Renal Context
Acute Kidney Injury In Cardiac Surgery Patients

Frederic Schneider

Final Presentation

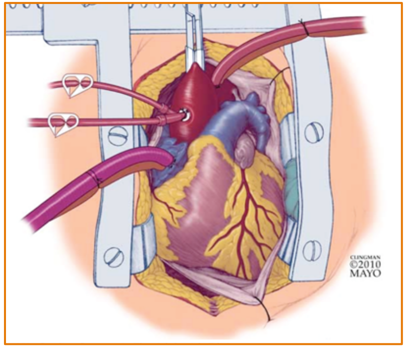
Trends in Bioinformatics, Winter 2017/18

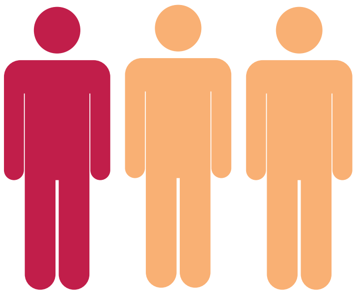
Acute Kidney Injury In Heart Patients (Recap)

- Heart and kidney are interconnected via various pathways
- Stress on either organ can cause dysfunction or injury of the other

- Cardiac Surgery is a common treatment for heart patients
  - Constitutes substantial stress for the heart

- Up to 30 % of patients undergoing cardiac surgery develop AKI [1]
- AKI is associated with substantial increase in morbidity and mortality





Previous Work On This Topic (Recap)

- Previous work focuses on detection of AKI onset
- Monitoring vitals and blood test results during ICU stay after surgery

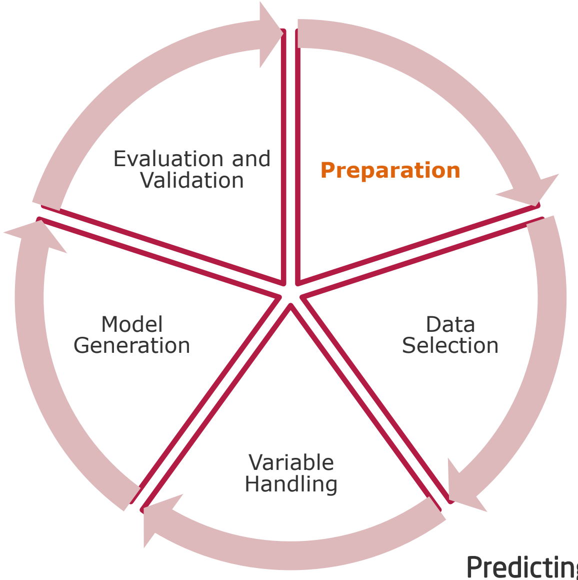
- Goal of this seminar work: Identifying patients who are at risk for AKI before surgery
- Analyzing patient records, laboratory values, patient data leading up to surgical intervention

Agenda

1. Motivation
2. Methods
   - Preparation & Data
   - **Model Generation**
   - **Local Interpretable Model-Agnostic Explanations (LIME)**
   - Proof-Of-Concept Architecture
3. **Preliminary Results**
4. Conclusion
5. Outlook

Methods
Preparation

- Target outcome to predict: Post-surgical acute kidney injury
- Target users: Clinical professionals

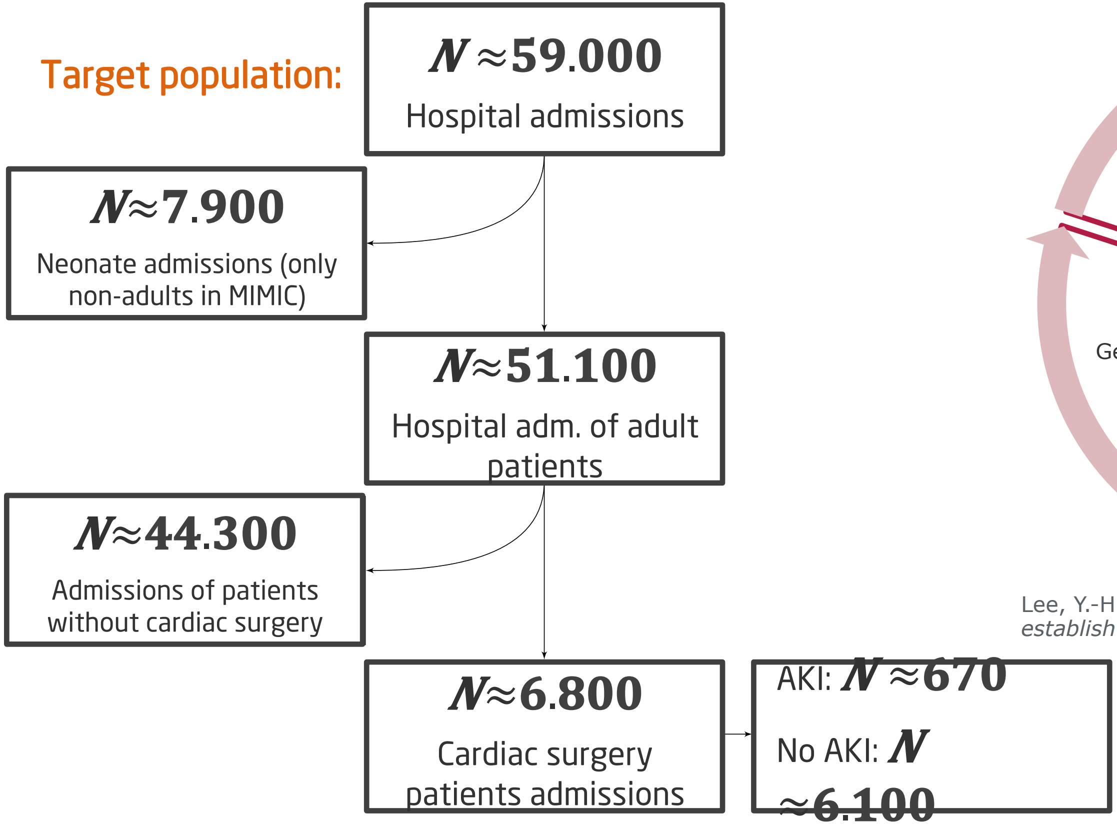


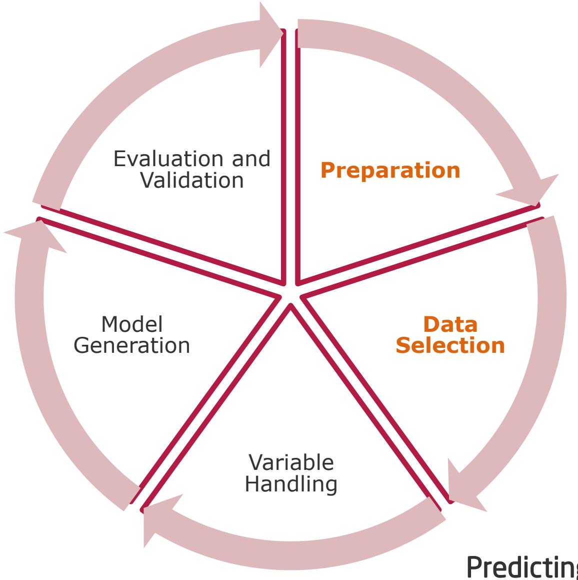
Lee, Y.-H., Bang H., & Kim, D.J., *How to establish clinical prediction models* (2016)

https://upload.wikimedia.org/wikipedia/commons/thumb/c/c3/Coronary_artery_bypass_surgery_Image_657C-PH.jpg/1200px-Coronary_artery_bypass_surgery_Image_657C-PH.jpg

Methods
Cohort Selection

- Target population:





Lee, Y.-H., Bang H., & Kim, D.J., *How to establish clinical prediction models* (2016)

Methods – Variable Handling
Features Selection

- Input Data of ~ **100 patient-related features:**
  - Demographics (age, sex, race, etc.)
  - Comorbidities
  - Laboratory test results

BLD_CREA_...	BLD_CREA_...	BLD_CREA_...	ANION_GA...	ANION_GA...	ANION_GA...
1.100	1.200	?	10	15	?
0.500	0.600	0.600	12	10	12
7.500	?	?	18	?	?
0.700	0.500	?	11	11	?
?	?	1.100	?	?	9
?	?	0.800	?	?	16
11.300	6.700	11.700	15	14	17
11.300	6.700	11.700	15	14	17
0.800	0.800	0.900	15	13	16

- Missing values imputation using mean imputation

- Expert input on laboratory test results: Importance of recent laboratory results and events:

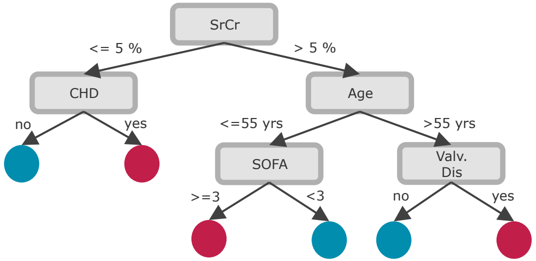
Methods – Variable Handling
Features Selection

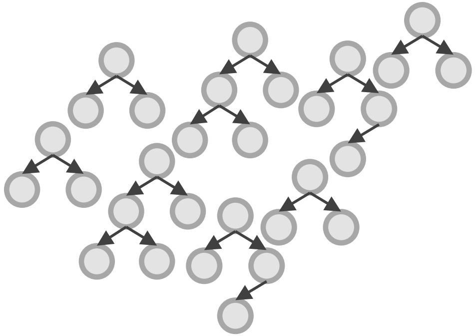
- Temporal context of laboratory results:
  - In MIMIC-III: Stored as event with timestamp
  - Feature extraction: Select relevant test results for 3 days before surgery and store as distinct features

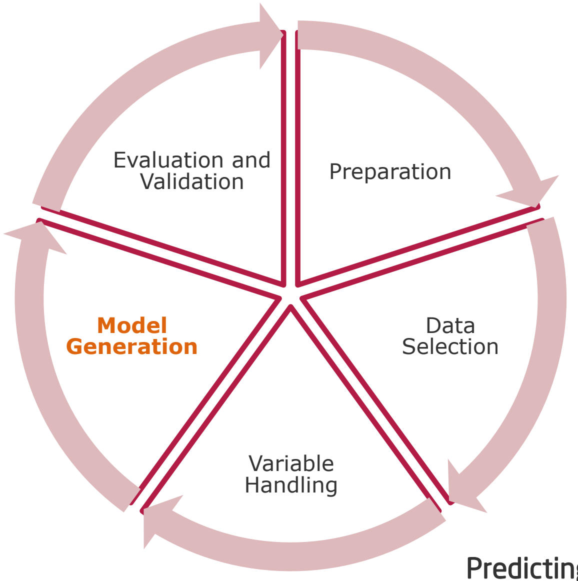
HADM_ID	CHARTTIME	VALUENUM	UOM	LABEL
130,079	Nov 3, 2182 5:05:00.0 PM	0.5	mg/dL	CREATININE
198,396	Aug 17, 2167 4:30:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 19, 2167 2:45:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 23, 2167 4:00:00.0 AM	0.5	mg/dL	CREATININE
130,079	Nov 13, 2182 4:21:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 23, 2167 3:00:00.0 PM	0.5	mg/dL	CREATININE
130,079	Nov 27, 2182 8:42:00.0 AM	0.5	mg/dL	CREATININE
130,079	Dec 2, 2182 6:55:00.0 AM	0.5	mg/dL	CREATININE
130,079	Dec 3, 2182 7:40:00.0 AM	0.5	mg/dL	CREATININE
136,012	Jan 15, 2145 5:45:00.0 AM	0.5	mg/dL	CREATININE
136,012	Jan 16, 2145 6:10:00.0 AM	0.5	mg/dL	CREATININE
136,012	Dec 31, 2144 4:00:00.0 AM	0.5	mg/dL	CREATININE
192,804	Jul 19, 2176 6:08:00.0 AM	0.5	mg/dL	CREATININE
192,804	Jul 20, 2176 5:10:00.0 AM	0.5	mg/dL	CREATININE
130,079	Dec 7, 2182 7:15:00.0 AM	0.5	mg/dL	CREATININE
130,079	Dec 8, 2182 7:45:00.0 AM	0.5	mg/dL	CREATININE
136,012	Jan 17, 2145 9:10:00.0 AM	0.5	mg/dL	CREATININE
136,012	Jan 1, 2145 1:30:00.0 AM	0.5	mg/dL	CREATININE
136,012	Dec 21, 2144 5:27:00.0 PM	0.5	mg/dL	CREATININE
130,079	Nov 2, 2182 2:49:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 16, 2167 5:00:00.0 PM	0.5	mg/dL	CREATININE
198,396	Aug 24, 2167 2:40:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 25, 2167 3:50:00.0 AM	0.5	mg/dL	CREATININE
198,396	Aug 28, 2167 4:30:00.0 AM	0.5	mg/dL	CREATININE
136,012	Dec 22, 2144 8:30:00.0 AM	0.5	mg/dL	CREATININE

BLD_CREA_...	BLD_CREA_...	BLD_CREA_...	ANION_GA...	ANION_GA...	ANION_GA...
1.100	1.200	?	10	15	?
0.500	0.600	0.600	12	10	12
7.500	?	?	18	?	?
0.700	0.500	?	11	11	?
?	?	1.100	?	?	9
?	?	0.800	?	?	16
11.300	6.700	11.700	15	14	17
11.300	6.700	11.700	15	14	17
0.800	0.800	0.900	15	13	16

Methods – Model Generation
Decision Trees And Gradient Boosted Decision Trees



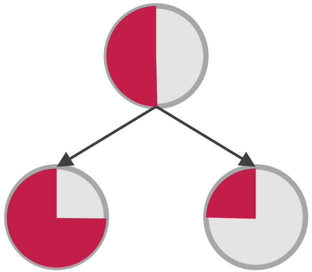




Lee, Y.-H., Bang H., & Kim, D.J., *How to establish clinical prediction models* (2016)

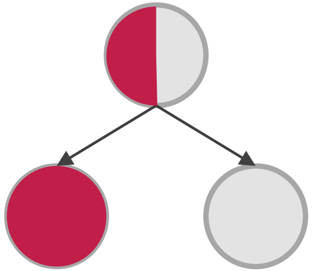
Methods
Decision Tree Training

- Given training set of size N, with feature vectors $x_i \in \mathbb{R}^d$, $i=1, \ldots, N$ and corresponding class assignments $c_1, \ldots, c_N$.



- Calculate for every new node in the tree the *best* split of the w.r.t. one of the d features at some threshold.

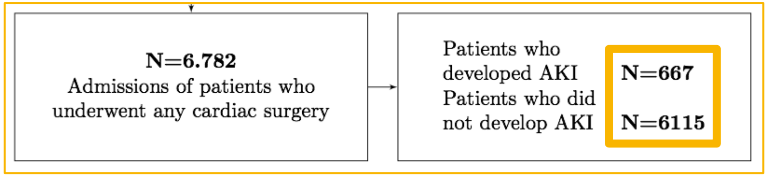
- Best split is determined using some impurity measure that should be minimal for the resulting split "populations" of training data, e.g. Gini-impurity: $Gini(E) = 1 - \sum_{i=1}^{|classes|} P_i^2$



Methods
Hyperparameter Tuning: Decision Trees

- Split criterion (Recap)
  - Gini impurity: $Gini(E)=1-\sum_{i=1}^{|classes|} P_i^2$
  - Information gain (entropy)

- Unbalanced training data
  - Class weight: $Gini(E)=1-\sum_{i=1}^{|classes|} w_i \cdot P_i^2$



→ Class weights AKI / no AKI: 10 / 1

- **Regularization hyperparameters:**
  - Minimum samples for split
  - Minimum impurity decrease
  - **Maximum tree depth**

Restrict growth of decision
→ Averts overfitting

- **Grid search** → Optimal performance on validation set using max. tree depth of 5

Methods
Gradient Boosted Decision Tree Training

- Uses residual fitting:

$F_m(x)=F_{m-1}(x)+\gamma_m h_m(x)$,

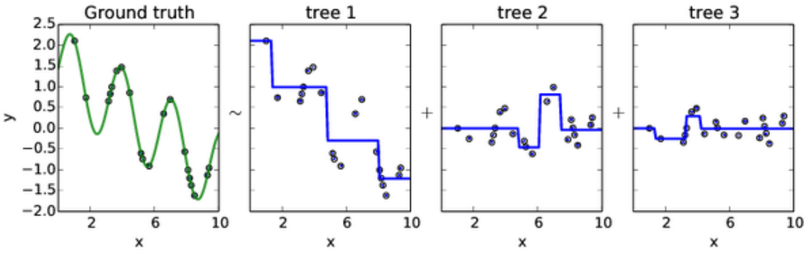
m is iteration,

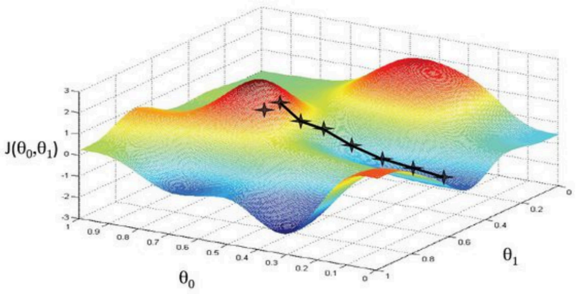
h_m is the m-th weak learner

- Employs decision trees as weak learners
- m-th weak learner is chosen so that it minimizes a loss function L, e.g. deviance for classification

$F_m(x)=F_{m-1}(x)-\arg\min_h \sum_{i=1}^{n} L(c_i, F_{m-1}+h_m(x_i))$

- **Steepest descent** is used to attempt to find **minimum of loss functions**
- The resulting trees then take a **weighted vote** to find a classification result





https://www.slideshare.net/sermakarevich/gradient-boosted-regression-trees-python

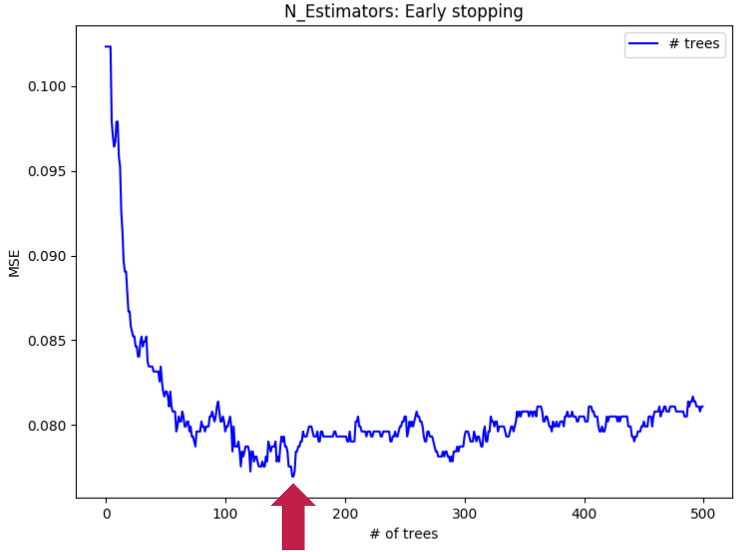
https://www2.eecs.berkeley.edu/Pubs/TechRpts/2015/EECS-2015-100.pdf

Methods
Hyperparameter Tuning: Gradient Boosted Decision Trees

- All hyperparameters of decision trees

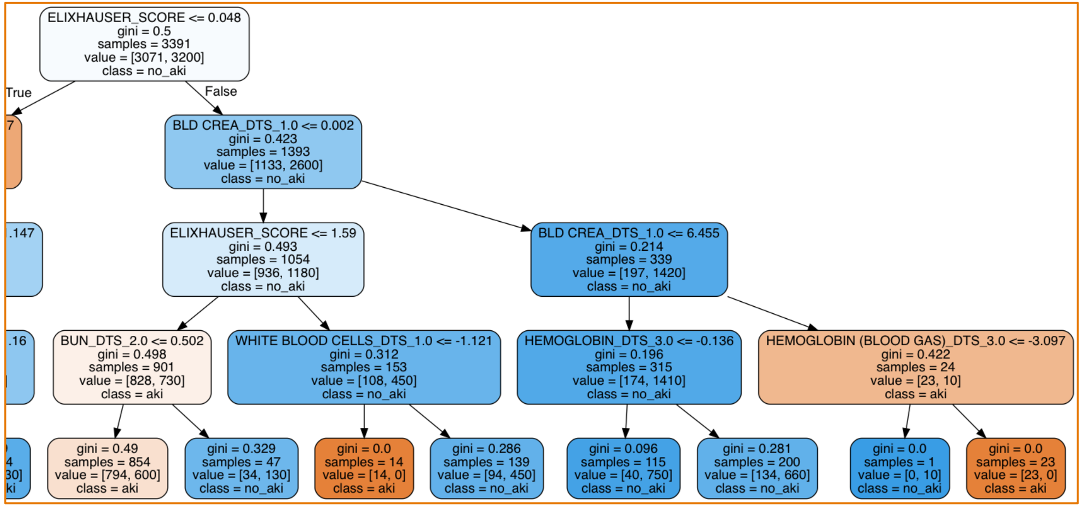
- Learning rate:
  - Contribution to result of each weak learner
  - Set to < 1, e.g. 0.1, → **Shrinkage** (regularization technique)
  - Low learning rate necessitates more learners but enables better generalization

- Number of weak learners (n_estimators in scikit-learn)
  - Can be determined using **early stopping** or target measure minimization



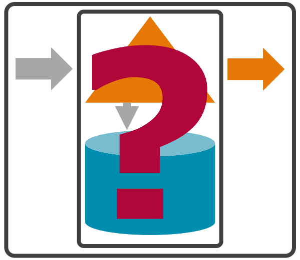
Methods
Prediction Model Interpretability

- **Decision Trees:** Model & decisions are **intelligible** to the human user



- **Gradient Boosted Decision Trees:** ensemble of > 100 trees

➡ **Neither model, nor specific results are intelligible**



Methods
Local Interpretable Model-Agnostic Explanations

- Uses easily explainable models, e.g. linear models, decision trees, rule lists
- Approximates non-interpretable models behavior at **position of explained results input**

Unfaithfulness of g to f in proximity to x

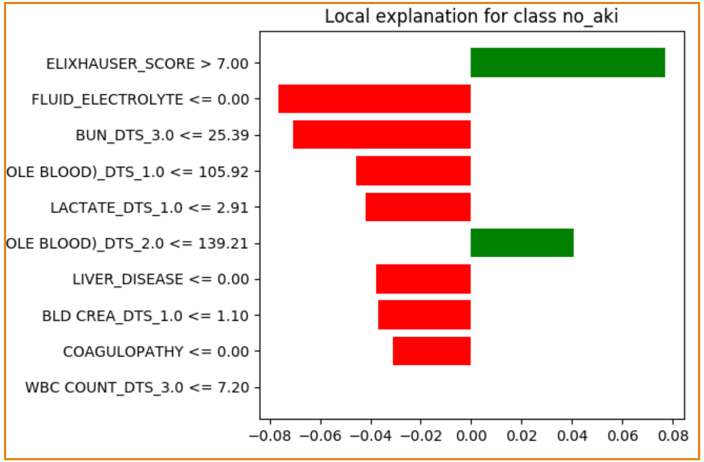
Actual model, explainable model, proximity measure for inputs other than x

- $\xi(x) = \arg\min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g)$

Explainable candidate models

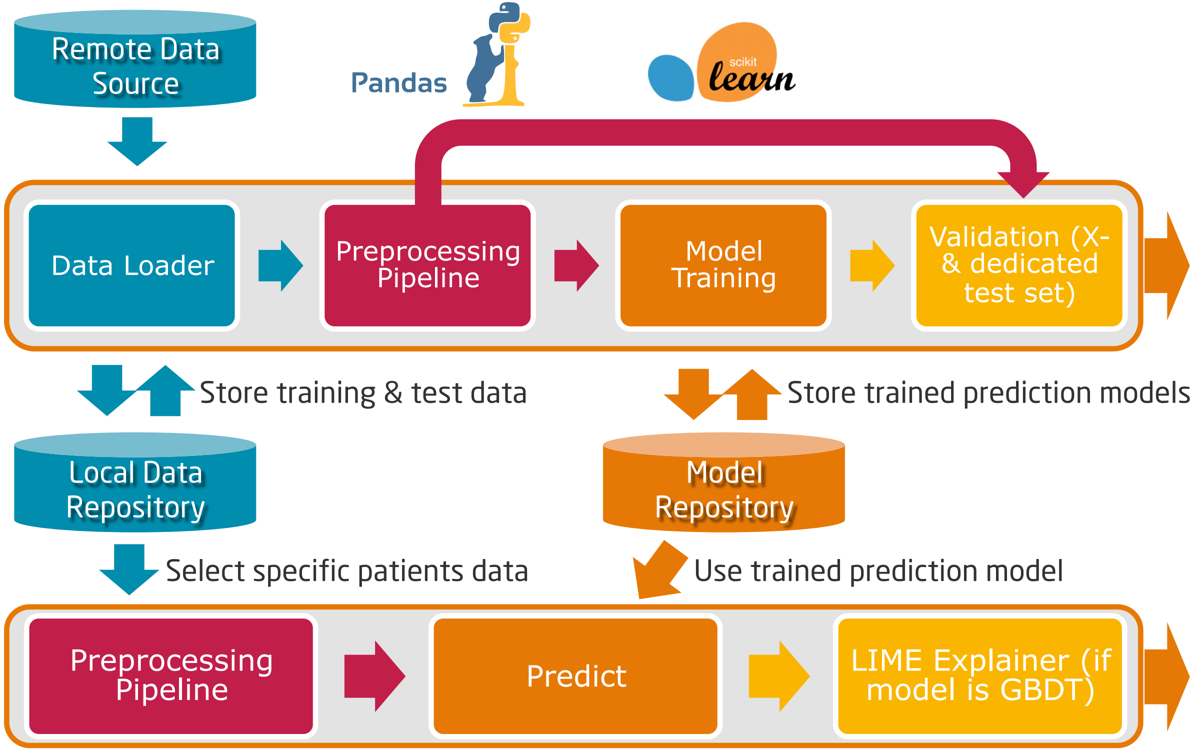
Penalty for complex models g

- Difference to mimic learning: Approximates only locally
- Explanation for classification result, not necessarily model



Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?": *Explaining the predictions of any classifier.* (2016)

Methods
Proof-Of-Concept Architecture



Performance Measures:

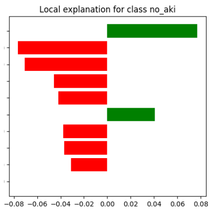
ROC curve, AUROC,

Confusion matrix,

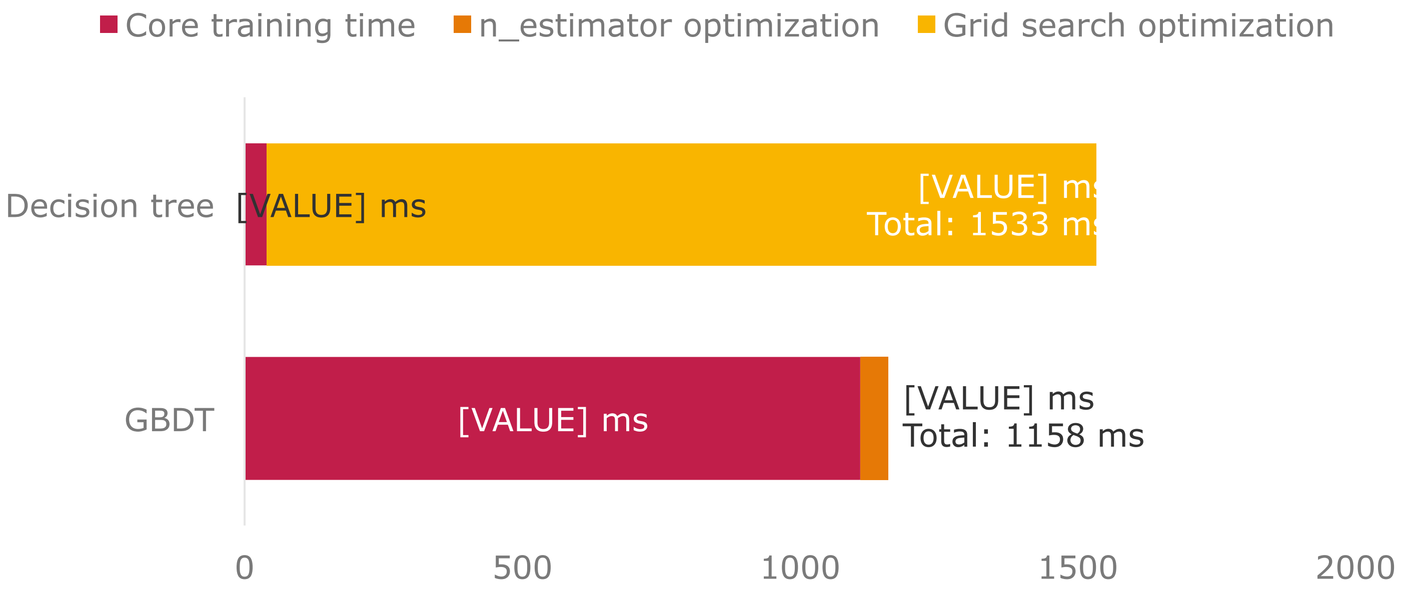
Recall, Accuracy

Prediction: AKI (Yes/No or likelihood)

Explanation:



Preliminary Results
Training Speed



Preliminary Results
Diagnostic Odds Ratio

Diagnostic Odds Ratio (DOR)

$$= True\ Positives / False\ Positives\ / False\ Negatives / True\ Negatives$$

- Measures the odds of a positive test result being correct relative to the probability of the test returning a false positive result

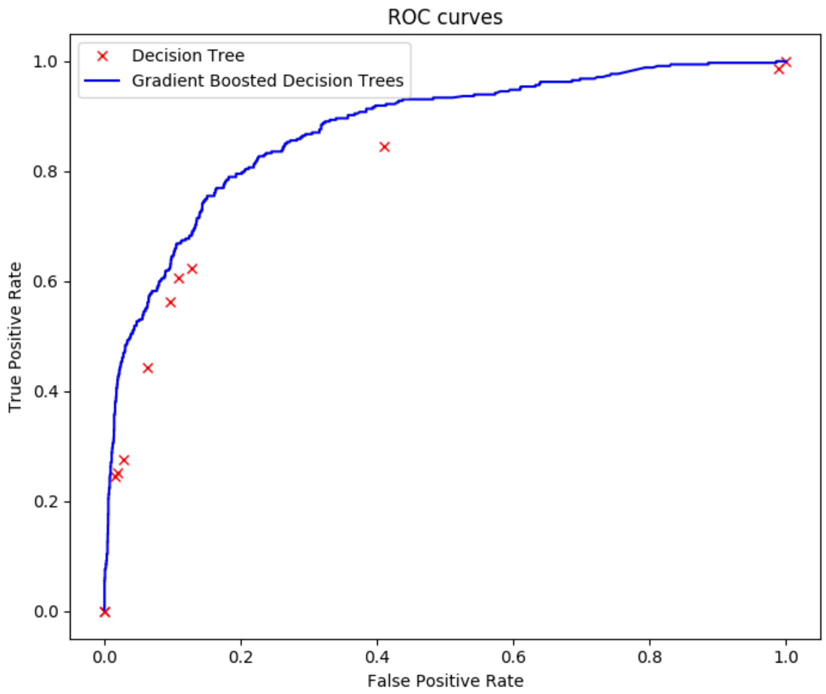
- Effectiveness measure for medical diagnostic tests
- Scalar value indicating test performance
- Independent of class distribution in test set

Preliminary Results
(Area Under) Receiver Operating Curve (AUROC)

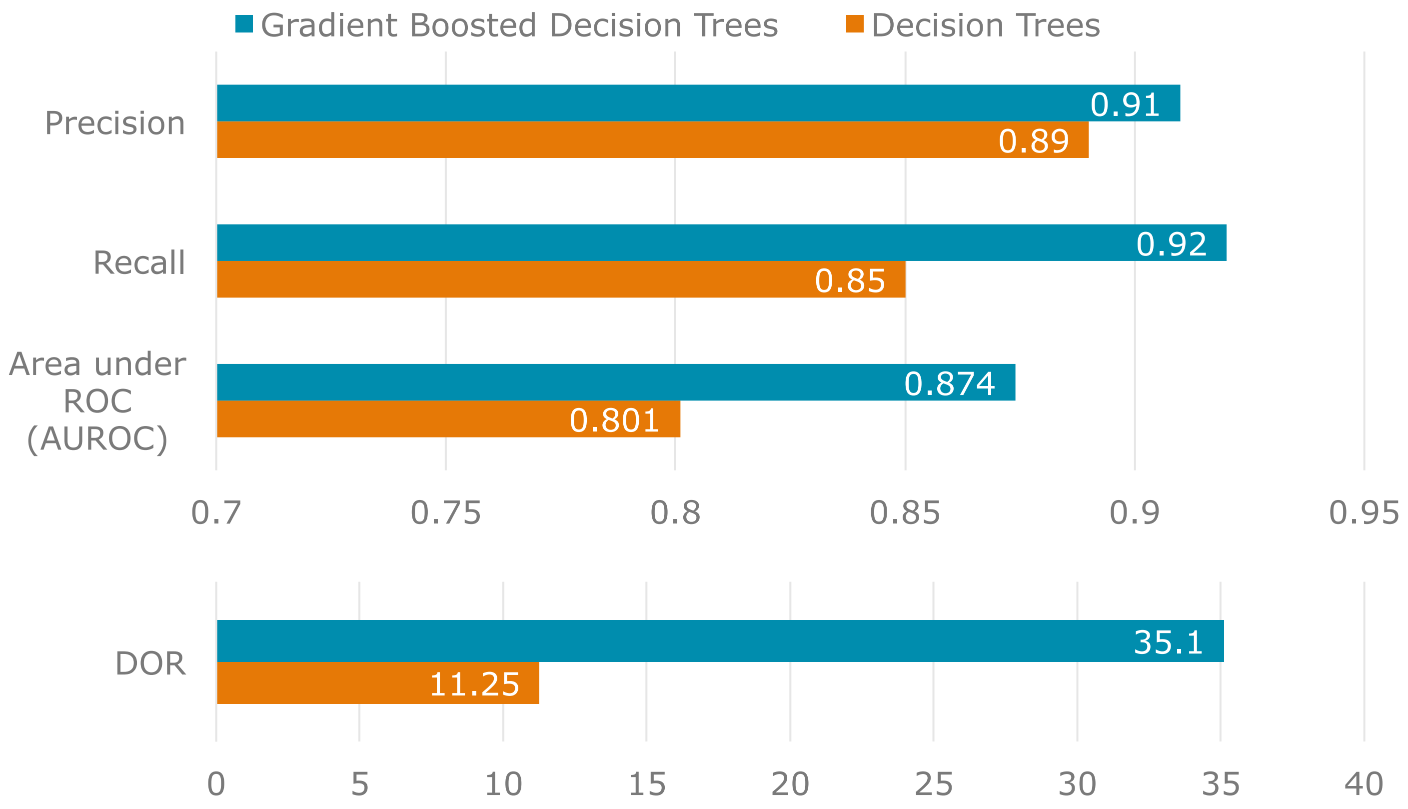
- Receiver operating curve plots **false positive rate** against **true positive rate**

- **AUROC** measures model fit regardless of accuracy/recall tradeoff

- **AUROC results:**
  - Decision trees: 0.801
  - Gradient boosted decision trees: 0.874

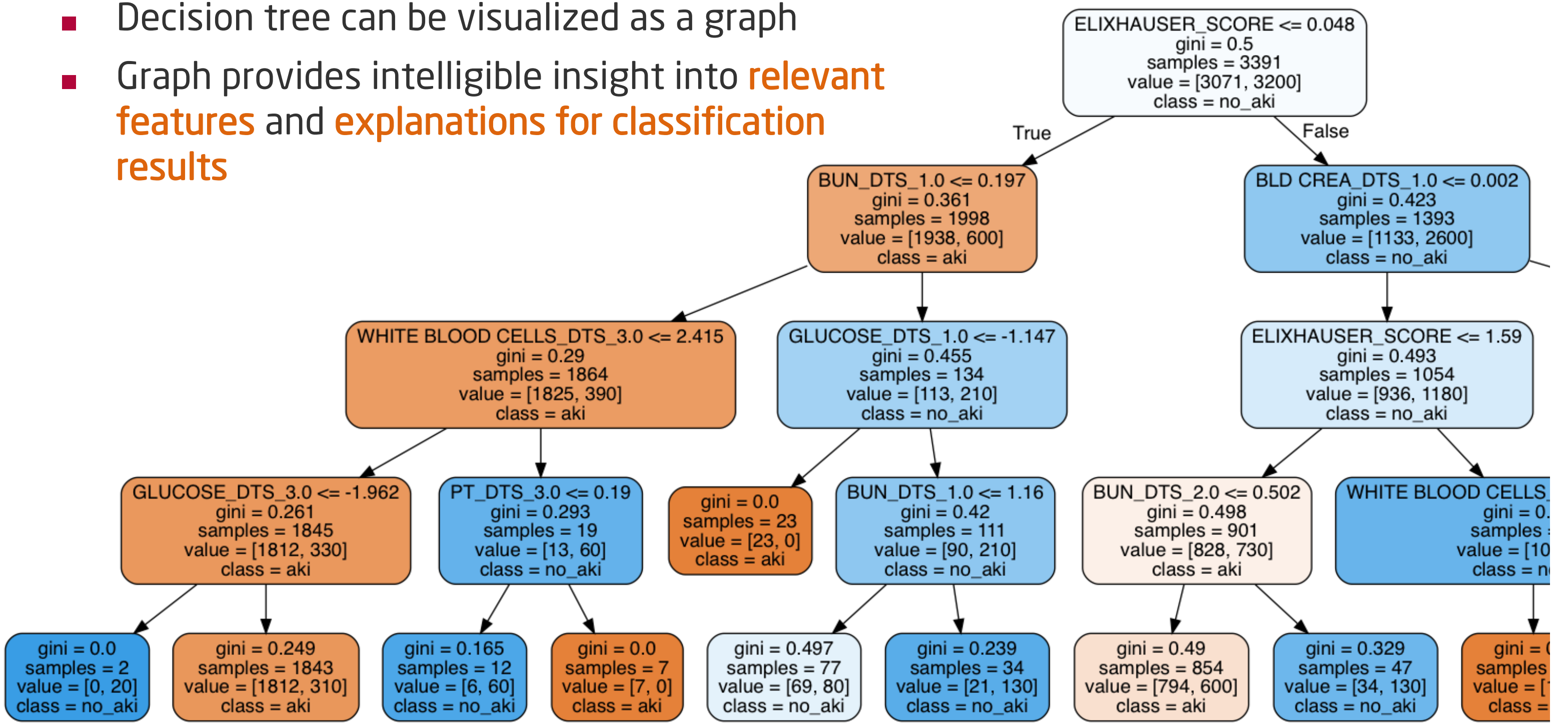


Preliminary Results
Precision, Recall, Area Under Receiver Operating Curve



Preliminary Results
Interpretability: Decision Tree

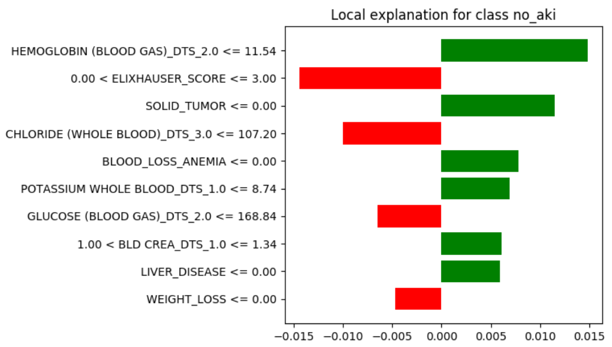
- Decision tree can be visualized as a graph
- Graph provides intelligible insight into **relevant features** and **explanations for classification results**

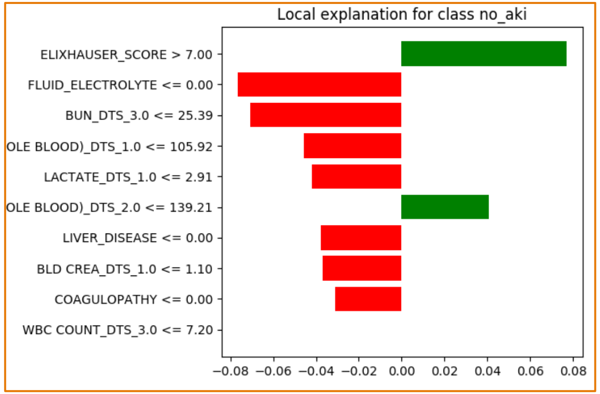


Visualization (excerpt) of trained decision tree model. Blue tint corresponds to low risk of AKI, orange to high risk of AKI

Preliminary Results
Interpretability GBDT With LIME

- GBDT do not provide a intelligible visualization of model behaviour
- **LIME offers human-interpretable insight on feature relevance for specific classification results**
- Recurring relevant features occur in decision tree as well:
  - Elixhauser score
  - Hemoglobin & Hematocrit
  - Creatinine levels
  - Glucose levels
  - Blood Urea Nitrogen

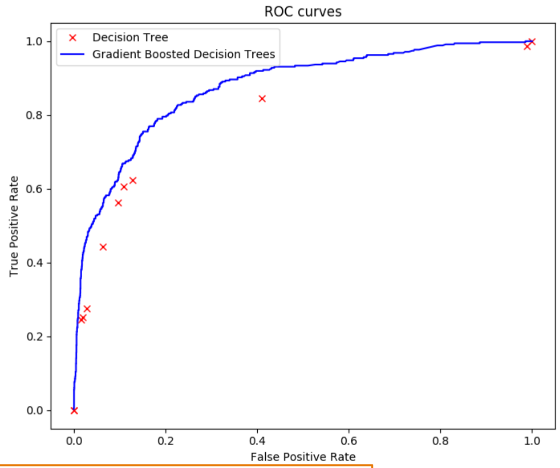




Conclusion

Developed POC application to predict risk of AKI in patients undergoing heart surgery

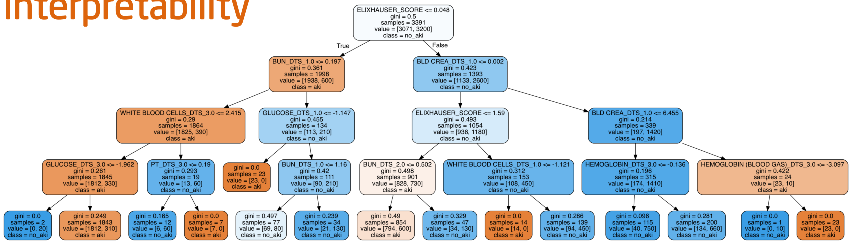
- Feature selection and preprocessing pipeline
- Compared two classifiers: Decision trees & GBDT
- Employs LIME to obtain result explanations

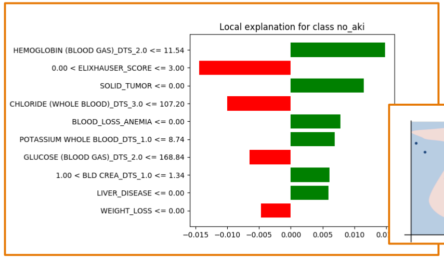


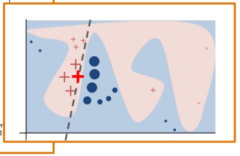
Performance advantage of GBDT

	Prec.	Rec.	DOR	AUROC
DT	0,89	0,85	11,25	0,801
GBDT	0,91	0,92	35,1	0,874

Interpretability





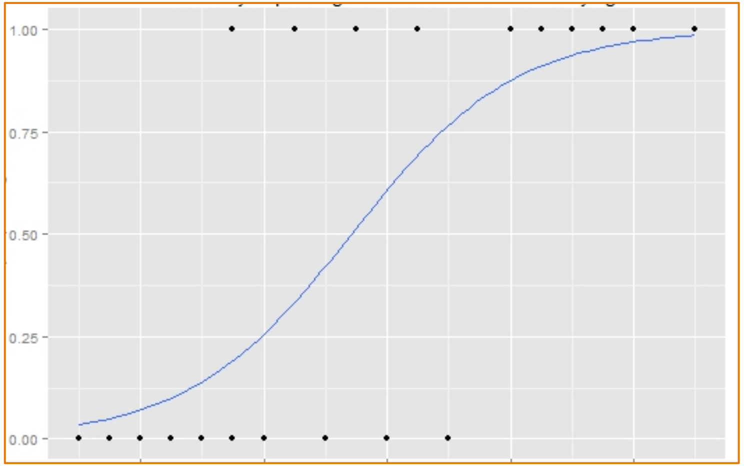


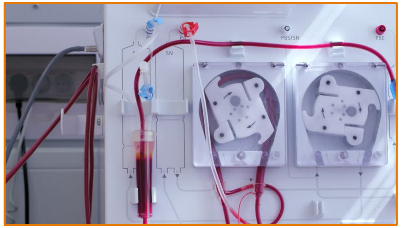
Outlook
Possible Improvements And Future Use-Cases

- **Different features**
  - More features, e.g. vitals

- Performance **comparison** with de-facto standard model in medicine: **Logistic Regression**

- **Predict different output variables**
  - 30 day and 90 day **mortality**
  - Readmission
  - Need for renal replacement therapy, i.e. **dialysis**





Q & A

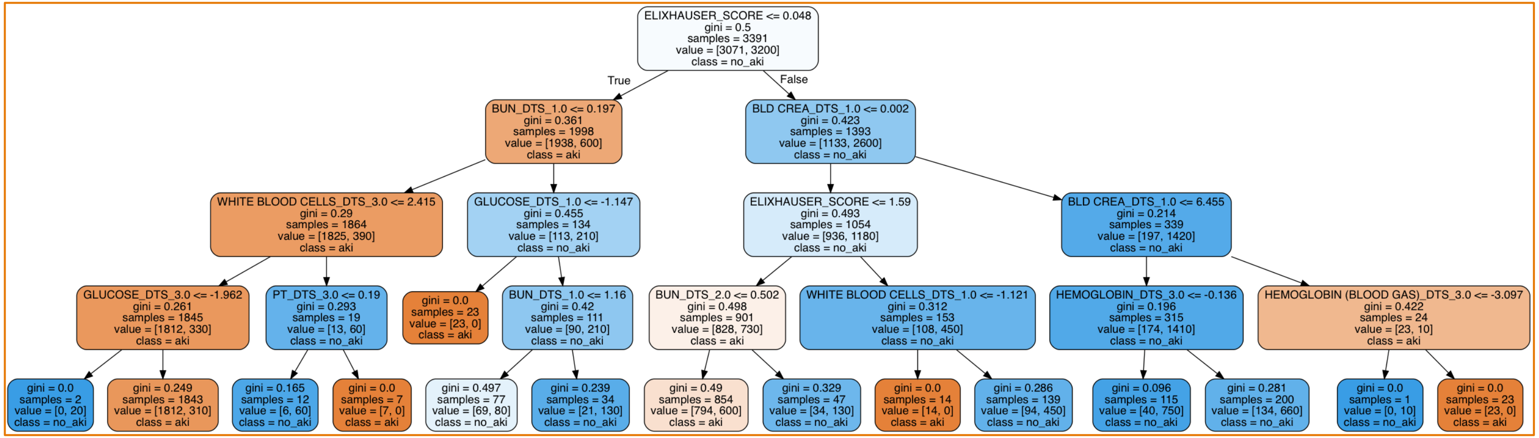
Any questions?

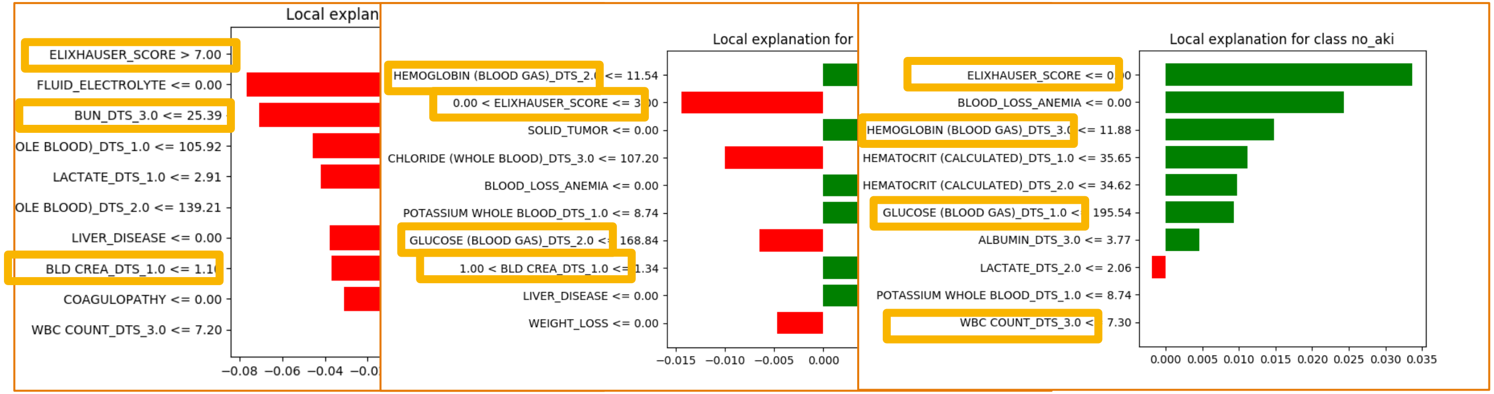
How would you handle missing data? Any ideas besides mean imputation?

Did the question of interpretability of machine learning techniques ever come up for you? When? Where?

What other patient data do you think could improve predictions?

Preliminary Results
Interpretability





Preliminary Results
Confusion Matrices, Precision, Recall, F1-score

Decision Trees

	Pred. No	Pred. AKI
Actual No	2655	389
Actual AKI	131	216

	Precision	Recall	F1	Support
No AKI	0.95	0.87	0.91	3044
AKI	0.36	0.62	0.45	347
Avg/total	0.89	0.85	0.86	3391

- Diagnostic Odds Ratio (DOR) ($True\ Positives/False\ Positives\ /False\ Negatives/True\ Negat$

Gradient Boosted Decision Trees

	Pred. No	Pred. AKI
Actual No	2991	53
Actual AKI	214	133

	Precision	Recall	F1	Support
No AKI	0.93	0.98	0.96	3044
AKI	0.72	0.38	0.50	347
Avg/total	0.91	0.92	0.91	3391

- DOR $\approx \mathbf{35{,}1}$

Thank you
for your attention!

Speaker

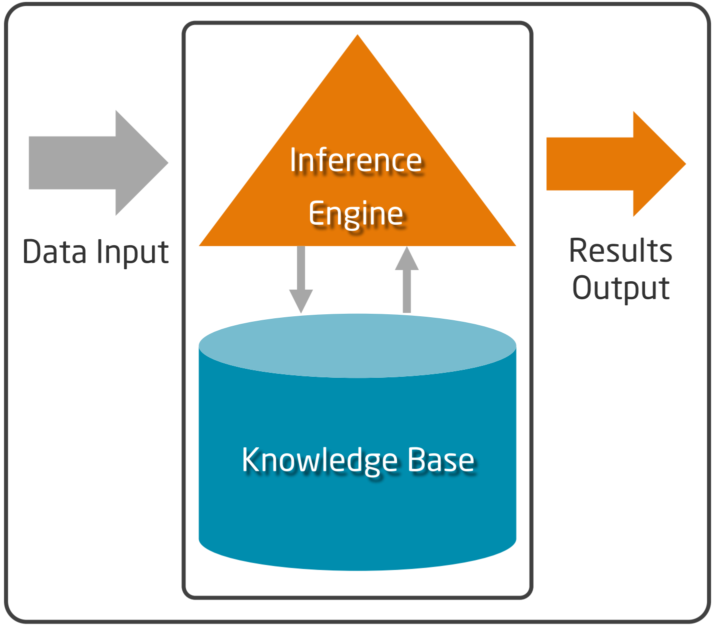
Job Description

Institute

Goal:
A Clinical Prediction Model For Post-Operative AKI

- Prediction of AKI before surgery
- Based on historical patient data
- Relevant outputs:
  - Risk for AKI
  - AKI stage
  - Confidence of classification
  - Need for renal replacement therapy

➡ Applicable in a clinical environment for decision support
  - How do you gain trust?
  - Does the result have an explanation?

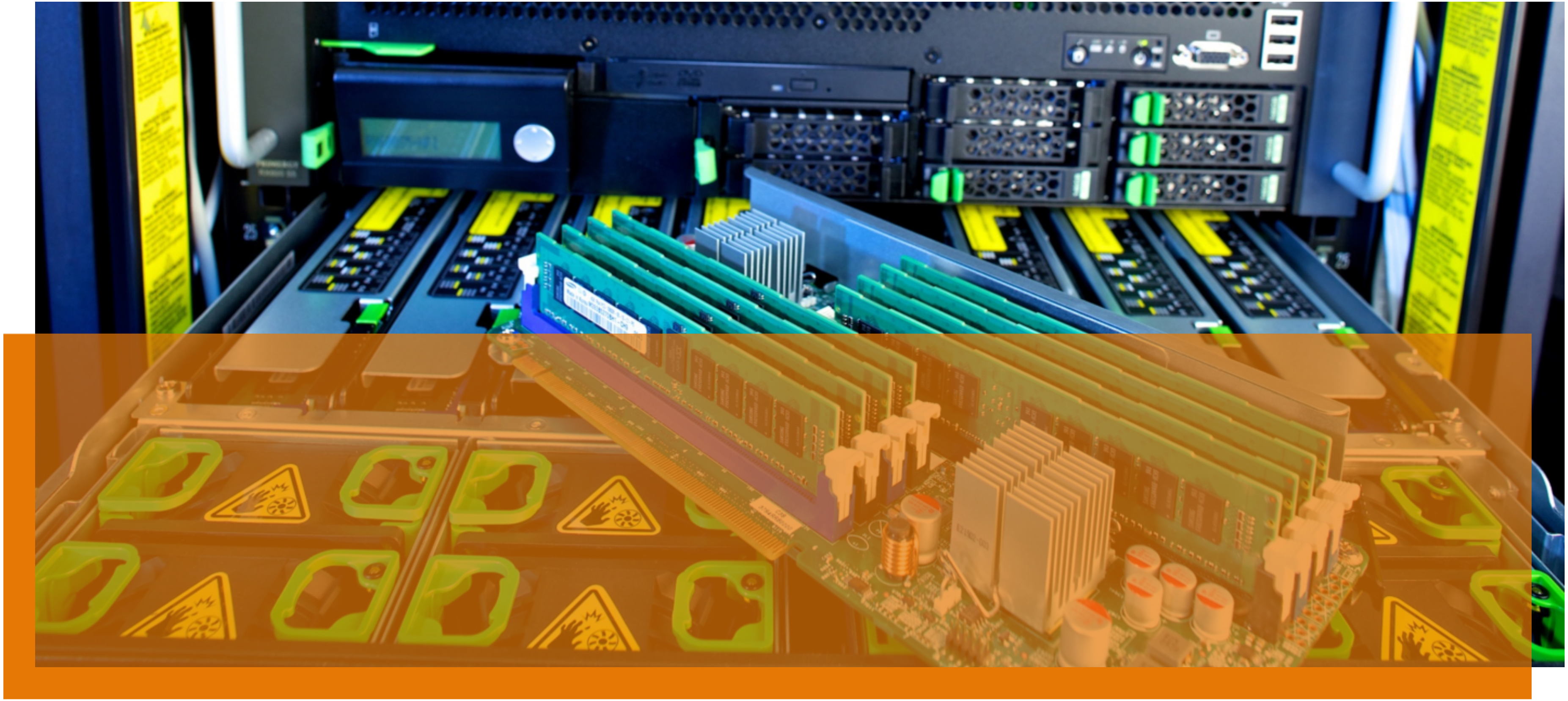


Architecture components of CDSS (Kola, n.d.)

Berner, E. (2009). Clinical decision support systems: state of the art. AHRQ Publication, (9) Retrieved from https://healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf

Universität Potsdam

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam



Universität Potsdam

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Universität Potsdam

HPI Hasso Plattner Institut
Digital Engineering • Universität Potsdam

Explanations
- Text layers

First text layer for running text.

- Second level for bullet points
  - Third level for bullet points
    - Fourth level for bullet points

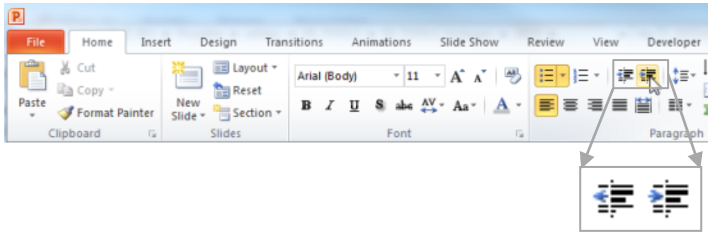
1. Fifth level for numberings
   a) Sixth level for listings

SEVENTH TEXT LAYER
FOR CORE MESSAGES

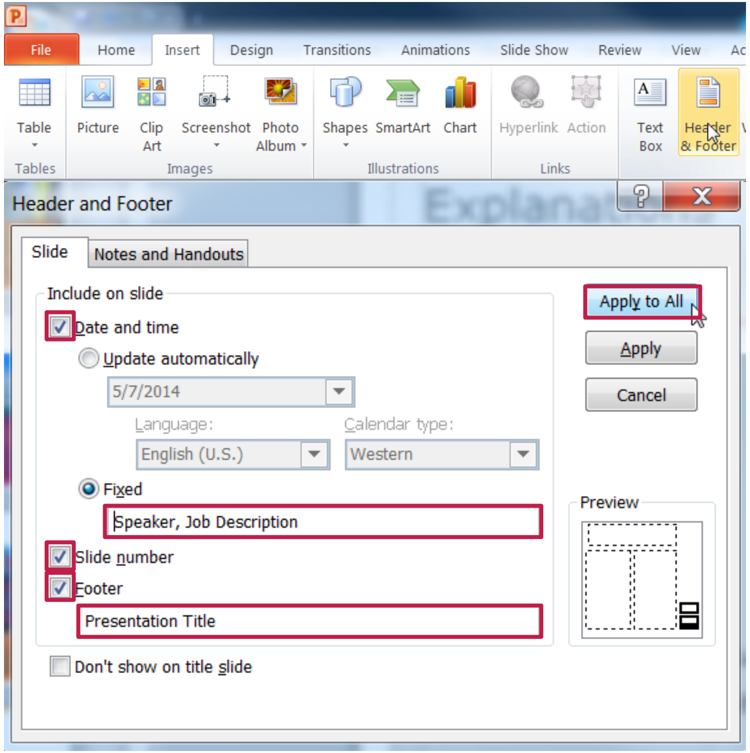
In this template, we pre-formatted different text layers
(as you can see on the right side).

You don’t have to generate bullet points manually.
By the way: Please avoid this!

To change from one text layer to the next, use the
Increase / Decrease List Level buttons:



Explanations
- Footer



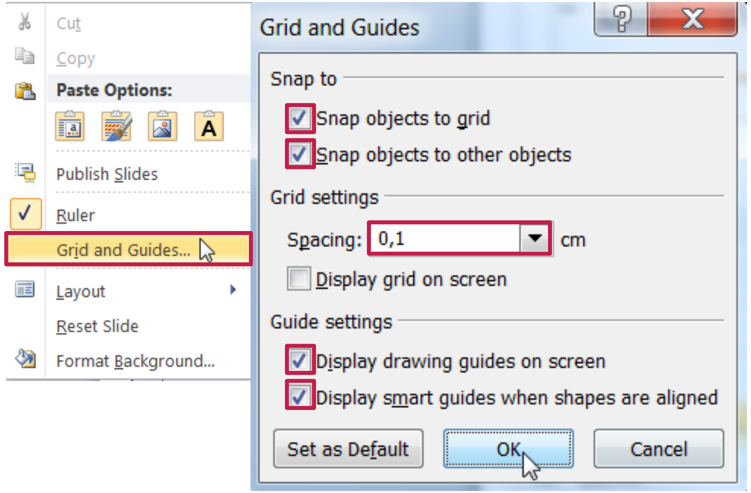
You can insert or change your presentation's footer. Click on the Insert-tab | Header and Footer | After filling in your descriptions click on **Apply to All.**

Descriptions:

- Activate date and time and write in:*Speaker, Job Description*
- Activate the slide number.
- Activate the footer and write in: *Presentation Title*

Don't use the template without the complete footer.

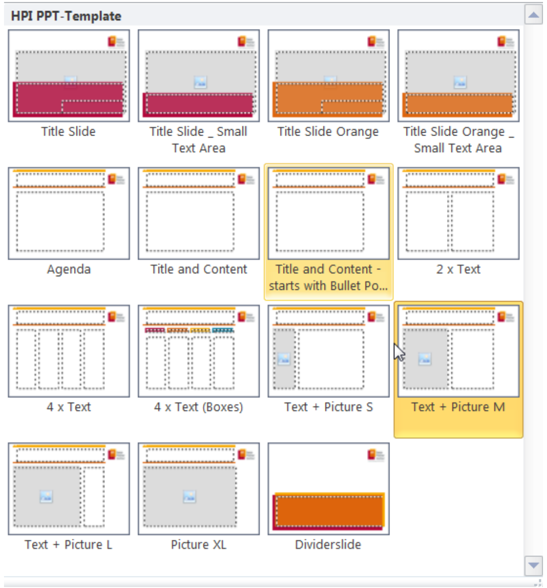
Explanations
- Drawing guides

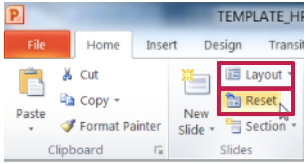


You can enable your guide-lines to align objects on the slide **(View | Show | Select the option „Guides“)**

Or hit the right mouse button outside the slide and go at „Grid and Guides…“

Explanations
- Slide layouts





You can choose between different slide layouts.
These pre-defined layouts gives you the oportunity to use text and visualisations just the right way.

To use these layouts:

Click on the Home-tab | New Slide or Layout | and choose one out of the layouts

Click „Reset“ to reset to the predefined slide layout.