FROM HEALTHY PEOPLE TO HEALTHY POWER PLANTS

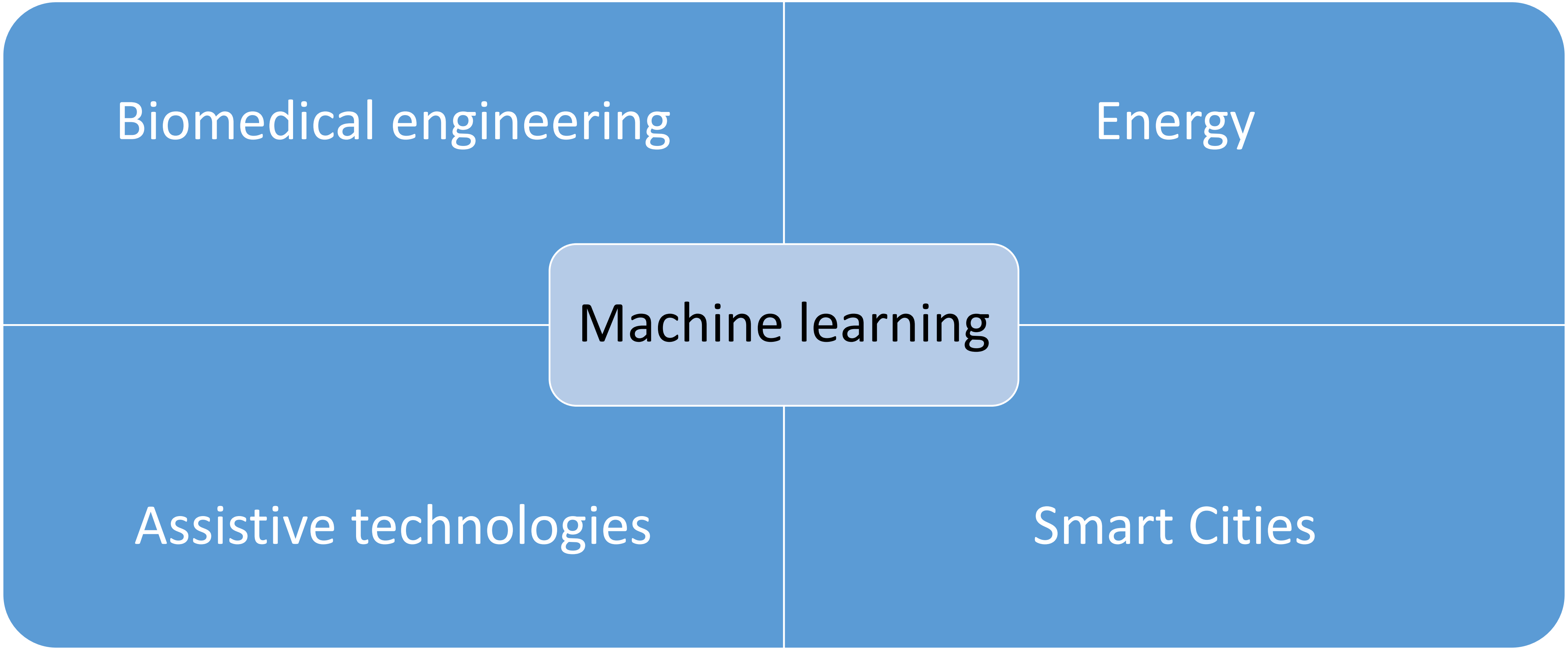
Martin Macas

Czech Institute of Informatics, Robotics and Cybernetics

Czech Technical University in Prague

martin.macas@cvut.cz

Introduction



Biomedical engineering

Energy

Machine learning

Assistive technologies

Smart Cities

Introduction – my research

Basic research

Swarm optimization, Feature selection, Multiple classifier system, Hidden Markov Models, Recurrent neural networks, Opinion formation models, Cluster analysis, Active machine learning, expert-in-the-loop classification, Multi-agent technologies

Biomedicine

Dyslexia detection from eye movements, EEG based emotion detection, Fetal heart rate signals classification, Cardiological signals processing, Mortality prediction, EEG based sleep staging, Sleep staging in neonatals, Intracranial pressure analysis, Clustering of EOG, Glycaemia prediction

Energy

Forecasting for HVAC systems, Heating control, Flexibility in energy consumption, Reliability of energy grid

Assistive technologies

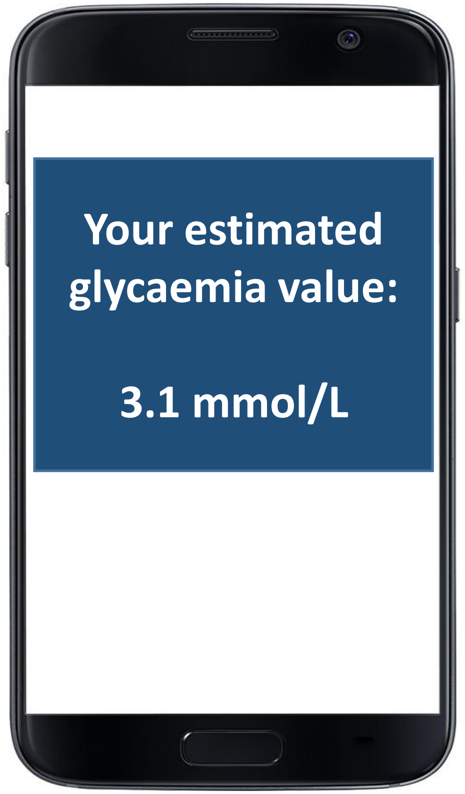
Camera based fall detection, Smart hospital bed

Other

Robot path planning, Electronic nose and tongue, Default prediction,

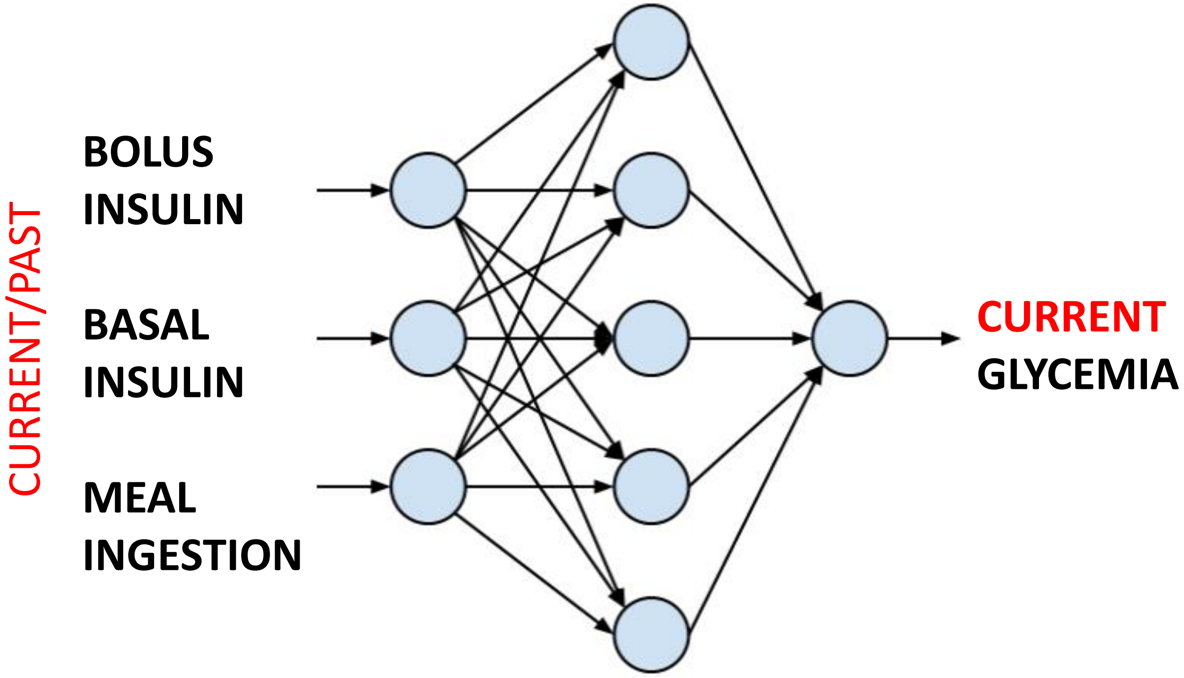
Example - ML for diabetics

- REGRESSION



Your estimated glycaemia value:

3.1 mmol/L



CURRENT/PAST

BOLUS INSULIN

BASAL INSULIN

MEAL INGESTION

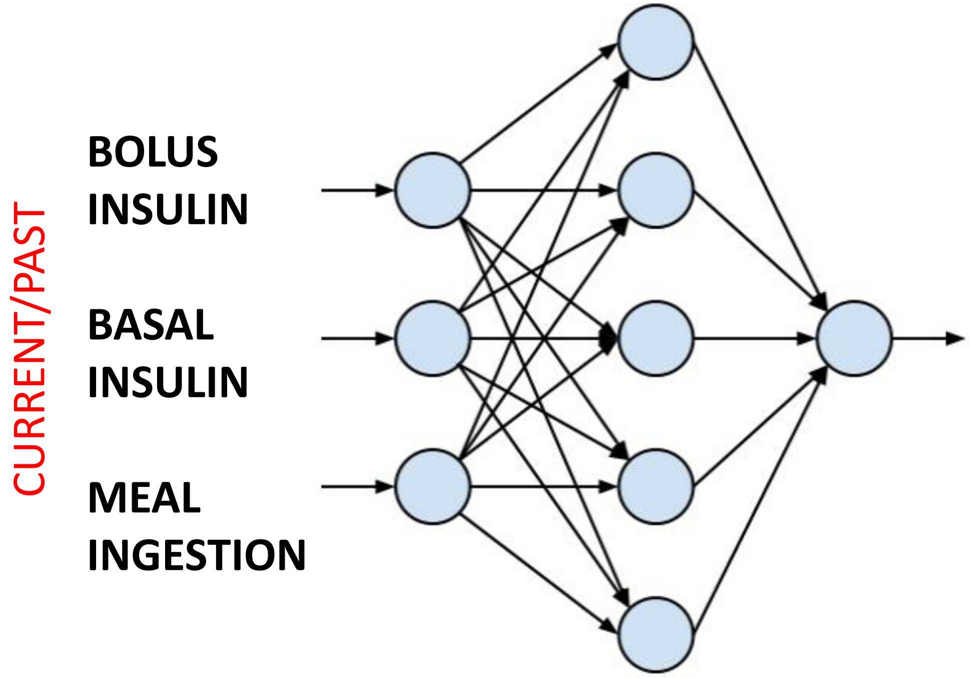
CURRENT GLYCEMIA

Example - ML for diabetics

- CLASSIFICATION

ALARM

YOU ARE NOW IN HYPOGLYCAEMIA !!!!!



CURRENT/PAST

BOLUS INSULIN

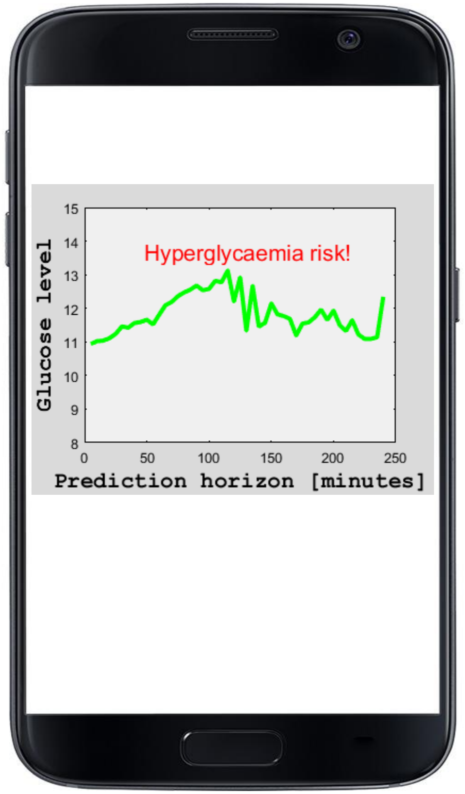
BASAL INSULIN

MEAL INGESTION

FLAG ABOUT CURRENT HYPER-GLYCEMIA (GLYCEMIA<3.9mmol/L)

Example - ML for diabetics

- FORECASTING

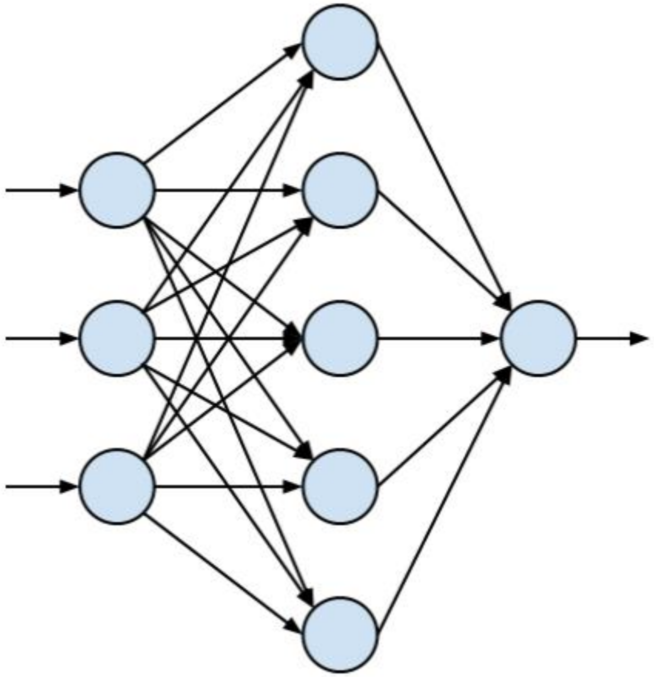


CURRENT/PAST/FUTURE

BOLUS INSULIN

BASAL INSULIN

MEAL INGESTION



FUTURE GLYCEMIA

Tasks in multidimensional time series processing

Regression

- Estimation of a relationship among independent variables and dependent variable

Classification

- Specific type of regression, where the dependent variable is categorical

Cluster analysis

- Like classification, but training data are not annotated – category of each training instance is not known

Regression/classification of sequential data

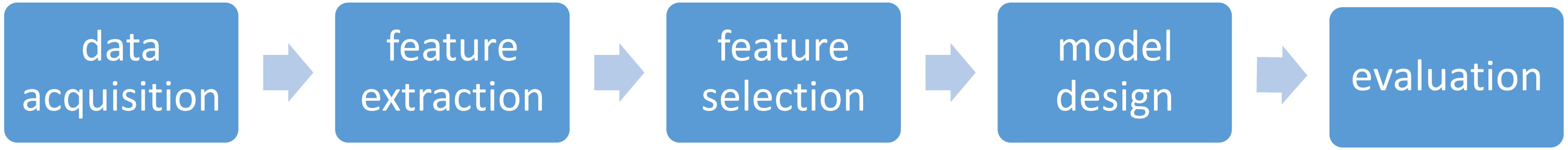
- Data instances are typically not “independent and identically distributed"

Forecasting

- Dependent variable is future value of some variable

Tasks in multidimensional time series processing

- Typical workflow



data acquisition

feature extraction

feature selection

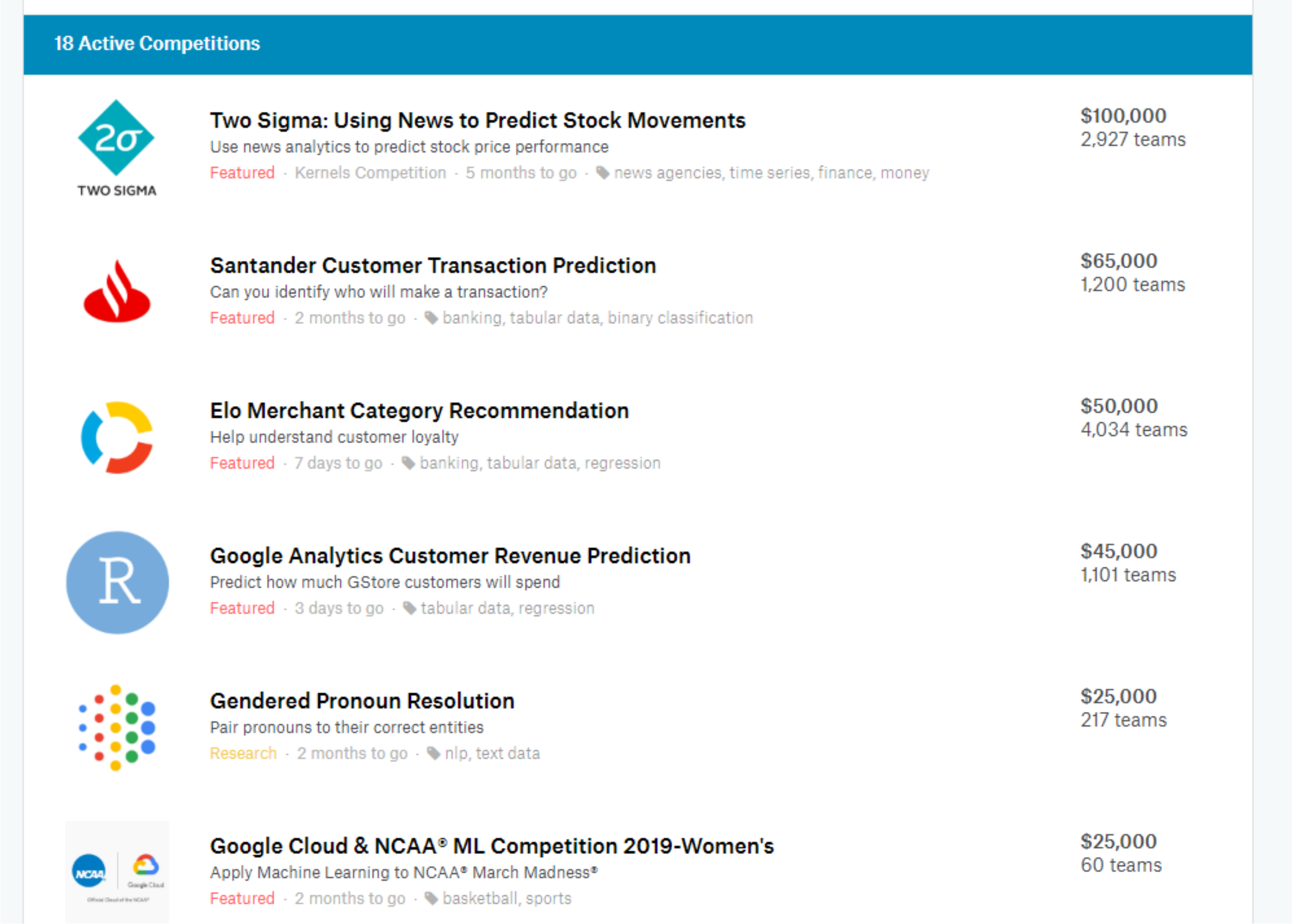
model design

evaluation

Lessons learned from competitions

- https://www.kaggle.com/

kaggle



18 Active Competitions

Two Sigma: Using News to Predict Stock Movements
Use news analytics to predict stock price performance
Featured · Kernels Competition · 5 months to go · news agencies, time series, finance, money
$100,000
2,927 teams

Santander Customer Transaction Prediction
Can you identify who will make a transaction?
Featured · 2 months to go · banking, tabular data, binary classification
$65,000
1,200 teams

Elo Merchant Category Recommendation
Help understand customer loyalty
Featured · 7 days to go · banking, tabular data, regression
$50,000
4,034 teams

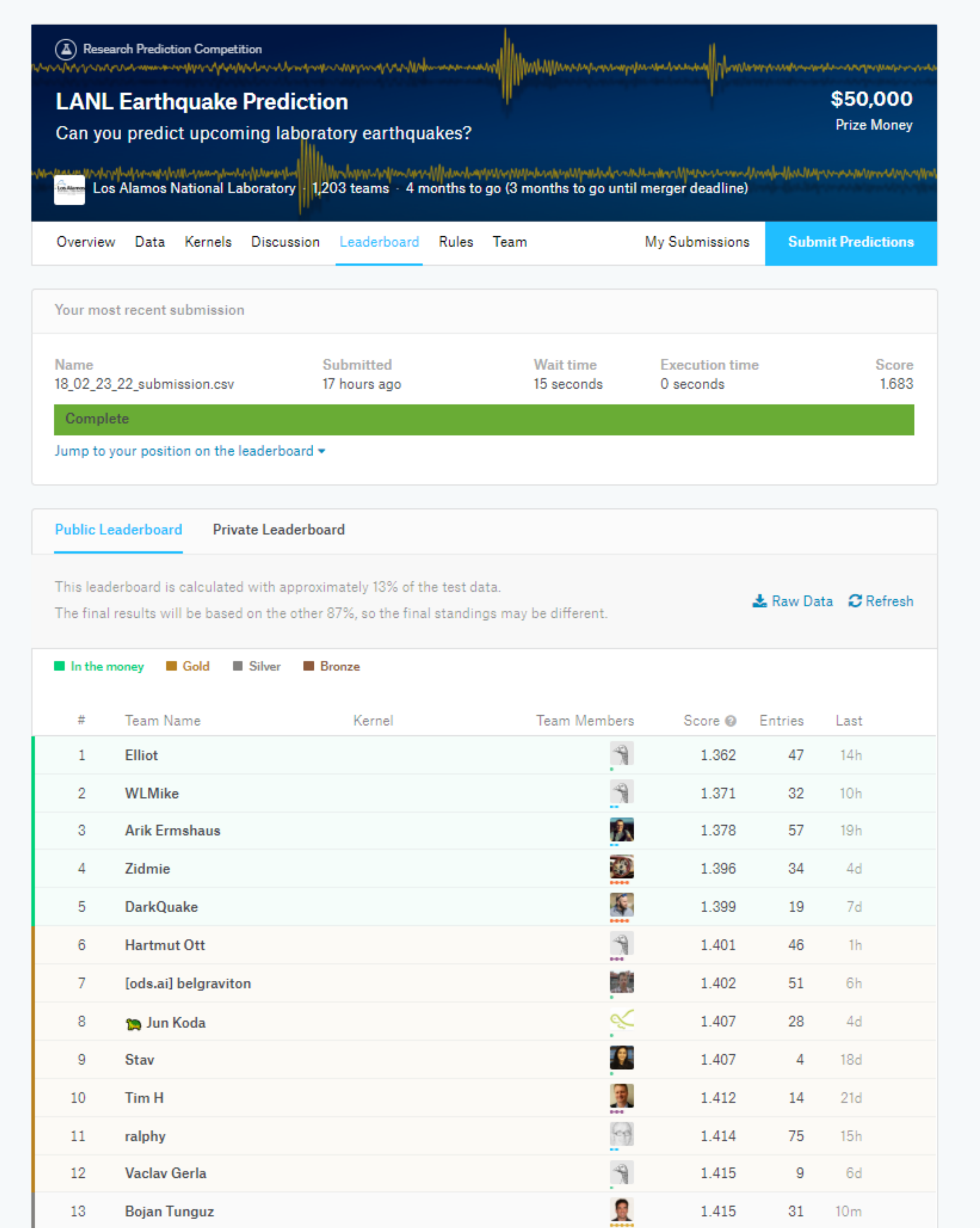
Google Analytics Customer Revenue Prediction
Predict how much GStore customers will spend
Featured · 3 days to go · tabular data, regression
$45,000
1,101 teams

Gendered Pronoun Resolution
Pair pronouns to their correct entities
Research · 2 months to go · nlp, text data
$25,000
217 teams

Google Cloud & NCAA® ML Competition 2019-Women's
Apply Machine Learning to NCAA® March Madness®
Featured · 2 months to go · basketball, sports
$25,000
60 teams

Lessons learned from competitions

- Participants register themselves and get
  - ANNOTATED training data that include input measurements and also target outputs (e.g. class labels)
  - UNANNOTATED testing data that include only input measurements and not target outputs
- Participants must submit predicted outputs
- Kaggle system compares target outputs with predicted outputs and computes an evaluation criterion (e.g. MSE, MAE, AUC …)
- Kaggle immediately provides feedback to participant
- Kaggle updates public leaderboard, where the participants can compare themselves with the others



Research Prediction Competition

LANL Earthquake Prediction

Can you predict upcoming laboratory earthquakes?

$50,000
Prize Money

Los Alamos National Laboratory · 1,203 teams · 4 months to go (3 months to go until merger deadline)

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
18_02_23_22_submission.csv	17 hours ago	15 seconds	0 seconds	1.683

Complete

Jump to your position on the leaderboard ▾

This leaderboard is calculated with approximately 13% of the test data.
The final results will be based on the other 87%, so the final standings may be different.

■ In the money ■ Gold ■ Silver ■ Bronze

#	Team Name	Kernel	Team Members	Score	Entries	Last
1	Elliot			1.362	47	14h
2	WLMike			1.371	32	10h
3	Arik Ermshaus			1.378	57	19h
4	Zidmie			1.396	34	4d
5	DarkQuake			1.399	19	7d
6	Hartmut Ott			1.401	46	1h
7	[ods.ai] belgraviton			1.402	51	6h
8	Jun Koda			1.407	28	4d
9	Stav			1.407	4	18d
10	Tim H			1.412	14	21d
11	ralphy			1.414	75	15h
12	Vaclav Gerla			1.415	9	6d
13	Bojan Tunguz			1.415	31	10m

Classification – mortality prediction

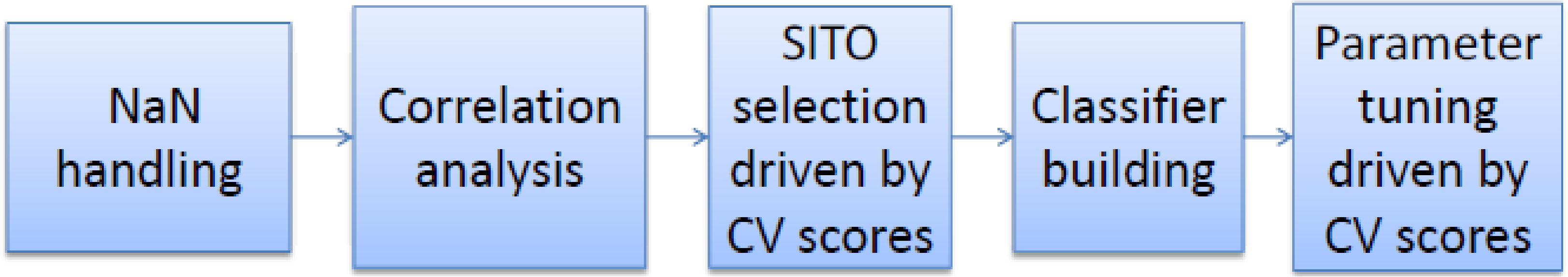
- Predicting Mortality of Intensive Care Unit Patients
- The PHYSIONET/COMPUTING IN CARDIOLOGY Challenge
- **2 evaluation criteria:**
  - Criterion 1: maximize min(sensitivity, positive predictivity)
  - Criterion 2: minimize Hosmer-Lemeshow statistic
  - 37 teams/participants
- **We achieved**
  - 4th place in Event 1
  - 3rd place in Event 2

Classification – mortality prediction

- Macaš, Martin, et al. "Linear Bayes classification for mortality prediction." *2012 Computing in Cardiology*. IEEE, 2012.
- Each record consisted of 37 time series of different lengths, each corresponding to one variable measured during the patient’s stay at ICU.
- The task was to predict if the patient will die earlier than one year after his stay at Intensive Care Unit – to detect “high-risk” patients

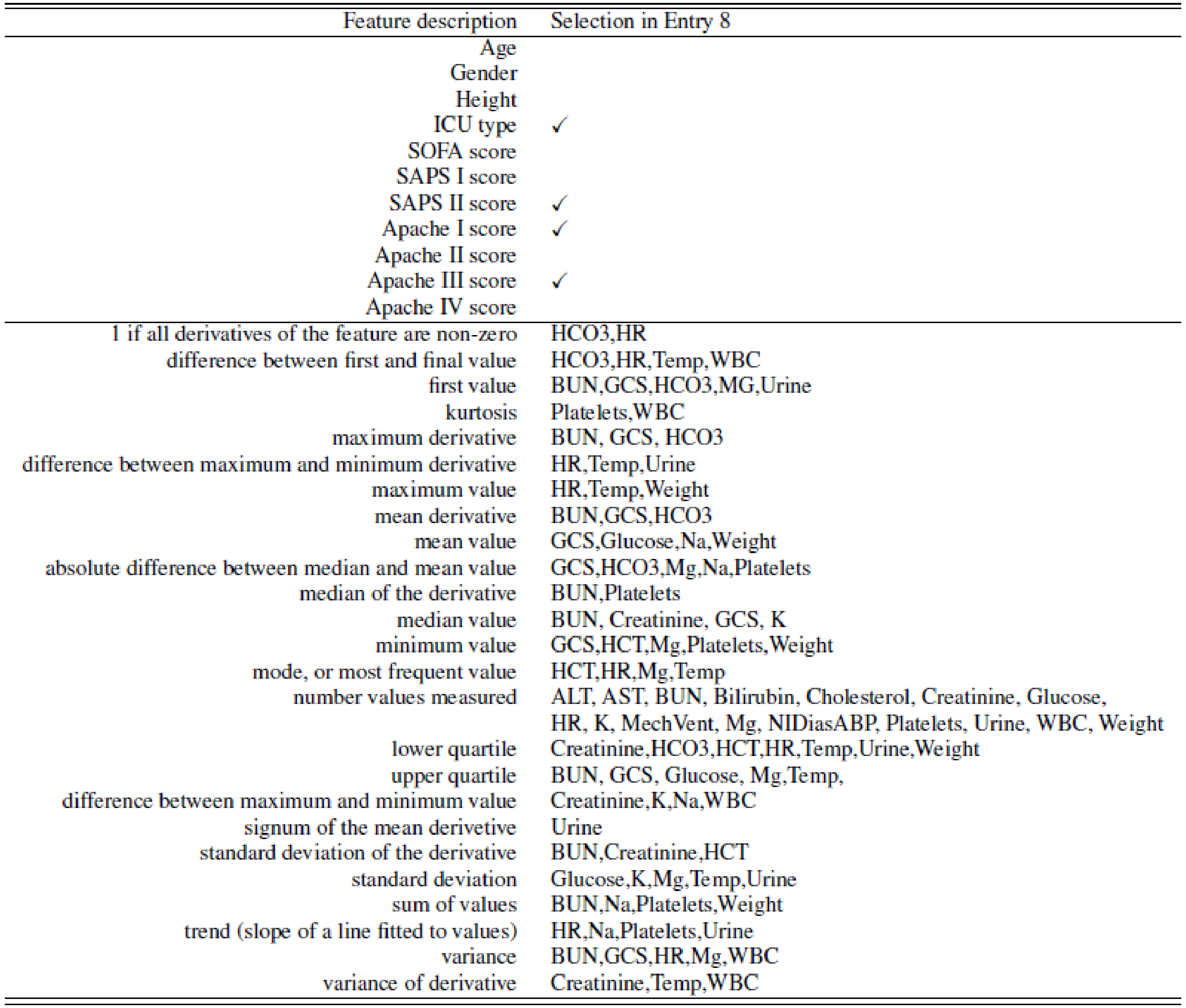
Classification – mortality prediction

- Our solutions:
  - Simple linear Bayes classifier
  - Great focus on feature extraction and feature subset selection
  - Using cross-validation to select features and optimize the system
  - Swarm intelligence method called Social Impact Theory based Optimization was used


NaN handling
Correlation analysis
SITO selection driven by CV scores
Classifier building
Parameter tuning driven by CV scores

Classification – mortality prediction

- 935 features extracted
- Feature selection performed



Feature description	Selection in Entry 8
Age	
Gender	
Height	
ICU type	✓
SOFA score	
SAPS I score	
SAPS II score	✓
Apache I score	✓
Apache II score	
Apache III score	✓
Apache IV score	
1 if all derivatives of the feature are non-zero	HCO3,HR
difference between first and final value	HCO3,HR,Temp,WBC
first value	BUN,GCS,HCO3,MG,Urine
kurtosis	Platelets,WBC
maximum derivative	BUN, GCS, HCO3
difference between maximum and minimum derivative	HR,Temp,Urine
maximum value	HR,Temp,Weight
mean derivative	BUN,GCS,HCO3
mean value	GCS,Glucose,Na,Weight
absolute difference between median and mean value	GCS,HCO3,Mg,Na,Platelets
median of the derivative	BUN,Platelets
median value	BUN, Creatinine, GCS, K
minimum value	GCS,HCT,Mg,Platelets,Weight
mode, or most frequent value	HCT,HR,Mg,Temp
number values measured	ALT, AST, BUN, Bilirubin, Cholesterol, Creatinine, Glucose, HR, K, MechVent, Mg, NIDiasABP, Platelets, Urine, WBC, Weight
lower quartile	Creatinine,HCO3,HCT,HR,Temp,Urine,Weight
upper quartile	BUN, GCS, Glucose, Mg,Temp,
difference between maximum and minimum value	Creatinine,K,Na,WBC
signum of the mean derivetive	Urine
standard deviation of the derivative	BUN,Creatinine,HCT
standard deviation	Glucose,K,Mg,Temp,Urine
sum of values	BUN,Na,Platelets,Weight
trend (slope of a line fitted to values)	HR,Na,Platelets,Urine
variance	BUN,GCS,HR,Mg,WBC
variance of derivative	Creatinine,Temp,WBC

Classification – mortality prediction

- Feature selection and parameter tuning based on maximization of cross-validation based criterion

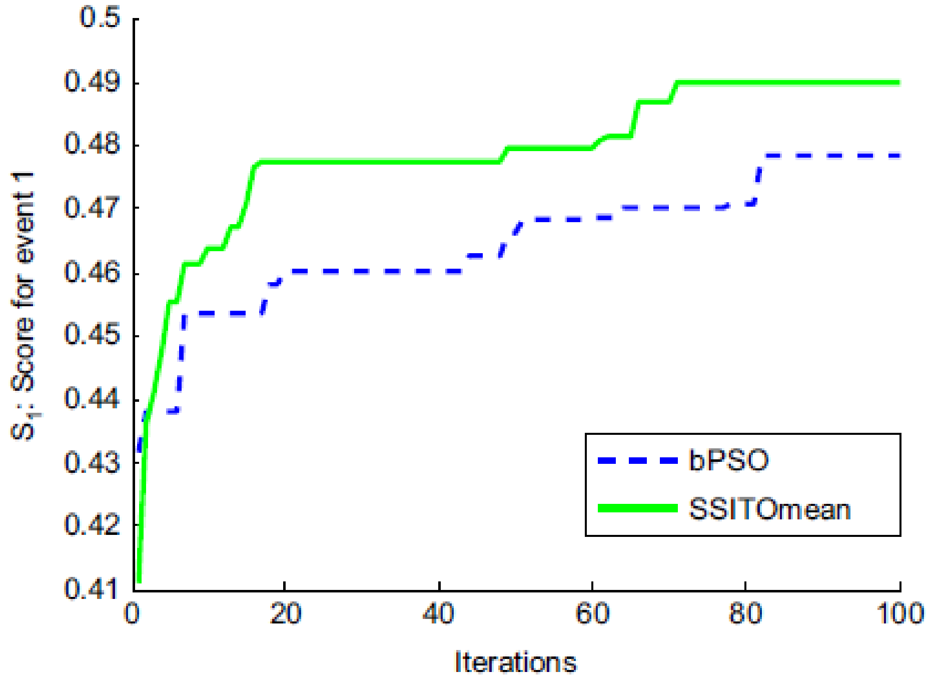
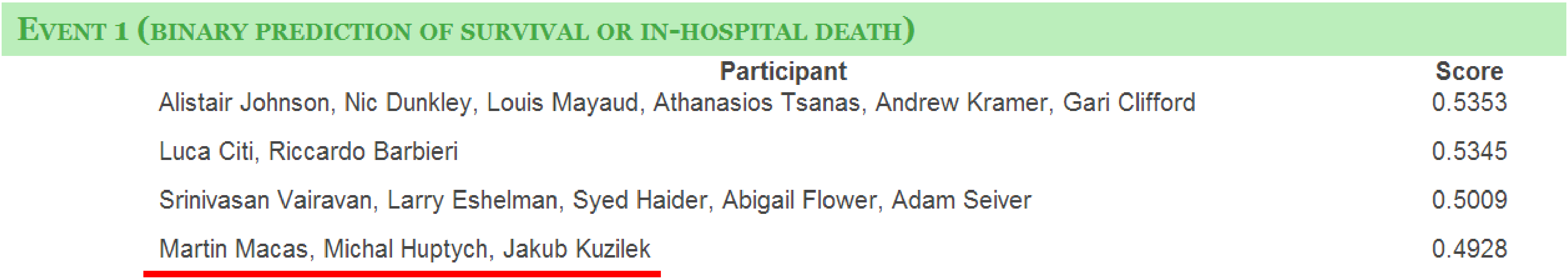


Fig. 5. Comparison of typical run of bPSO and SSITOmean algorithms on maximization of cross-validated score for event 1 in the CINC/PhysioNet Challenge.

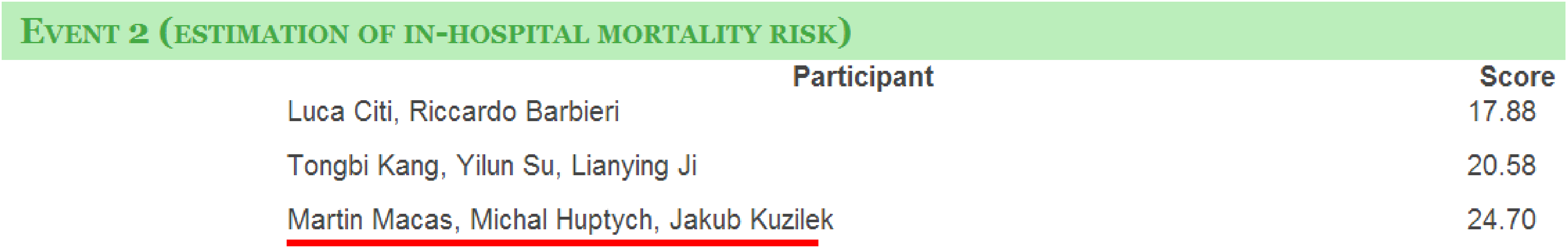
Classification – mortality prediction

- Winner’s solutions:
  - Ensemble of six support vector machines whose outputs were combined via regression



EVENT 1 (BINARY PREDICTION OF SURVIVAL OR IN-HOSPITAL DEATH)

Participant	Score
Alistair Johnson, Nic Dunkley, Louis Mayaud, Athanasios Tsanas, Andrew Kramer, Gari Clifford	0.5353
Luca Citi, Riccardo Barbieri	0.5345
Srinivasan Vairavan, Larry Eshelman, Syed Haider, Abigail Flower, Adam Seiver	0.5009
Martin Macas, Michal Huptych, Jakub Kuzilek	0.4928

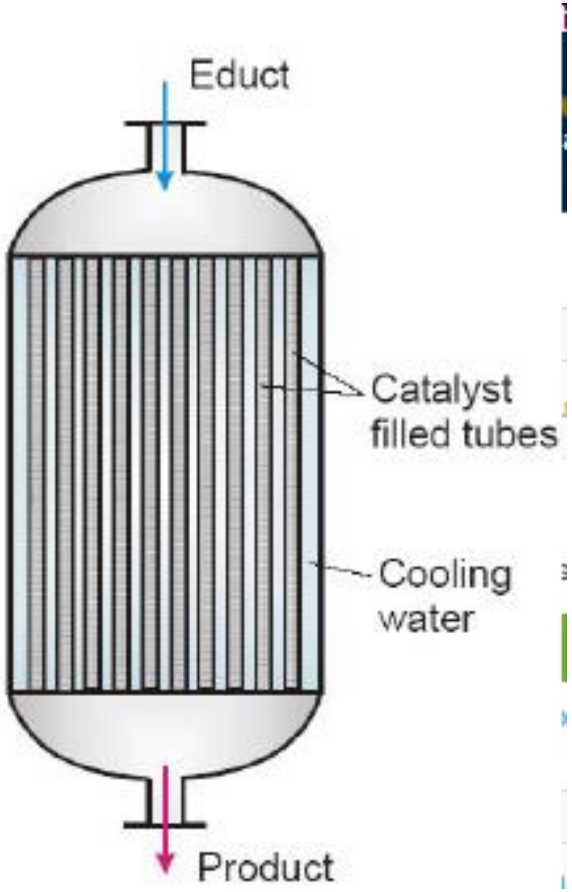


EVENT 2 (ESTIMATION OF IN-HOSPITAL MORTALITY RISK)

Participant	Score
Luca Citi, Riccardo Barbieri	17.88
Tongbi Kang, Yilun Su, Lianying Ji	20.58
Martin Macas, Michal Huptych, Jakub Kuzilek	24.70

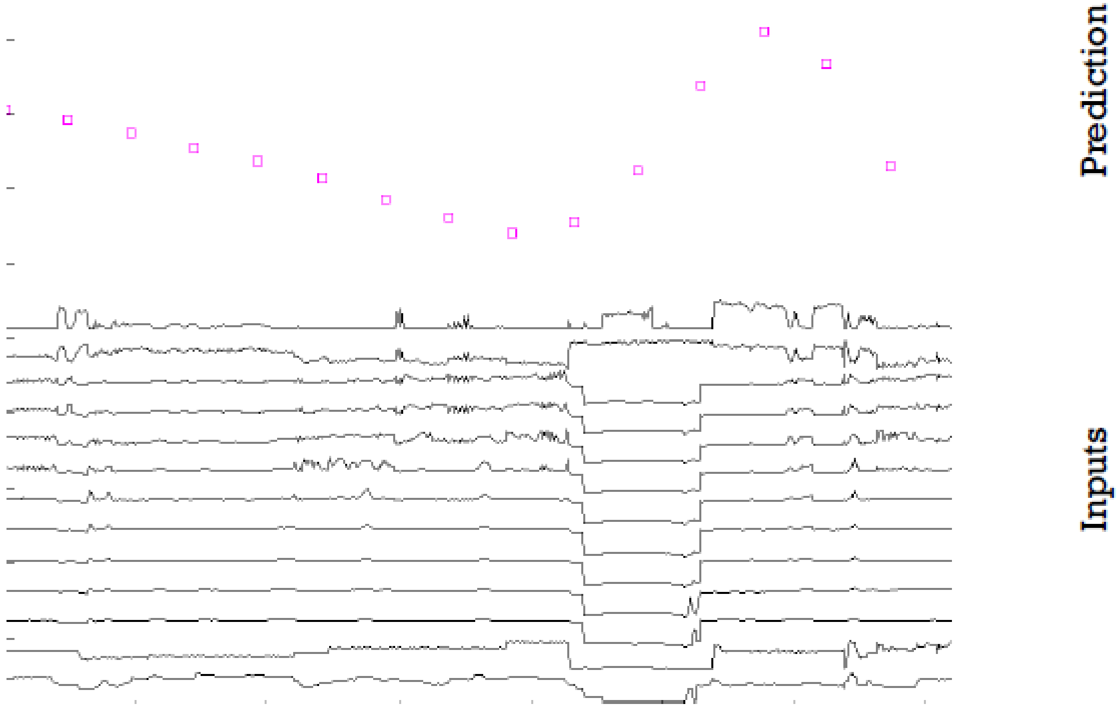
Prediction – chemical reactor activity

- The goal: to create an ADAPTIVE predictor that can adapt on changes caused by unmeasurable influences
- Inputs: 17 input variables
- Dependent variable: future activity



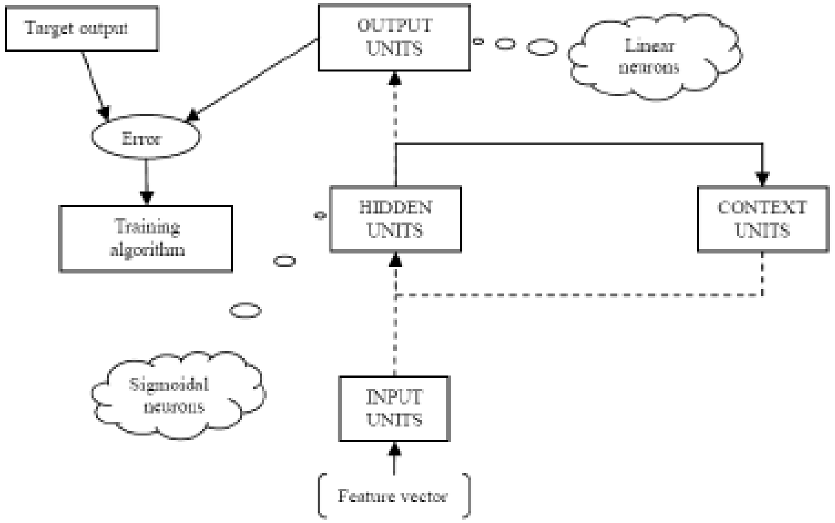
Prediction – chemical reactor activity

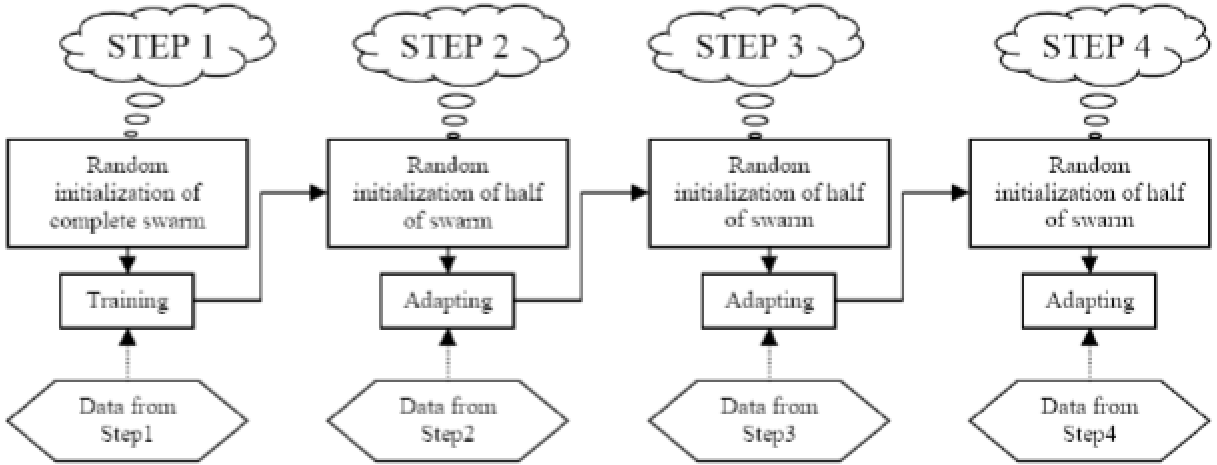
- Example of inputs and prediction



Prediction – chemical reactor activity

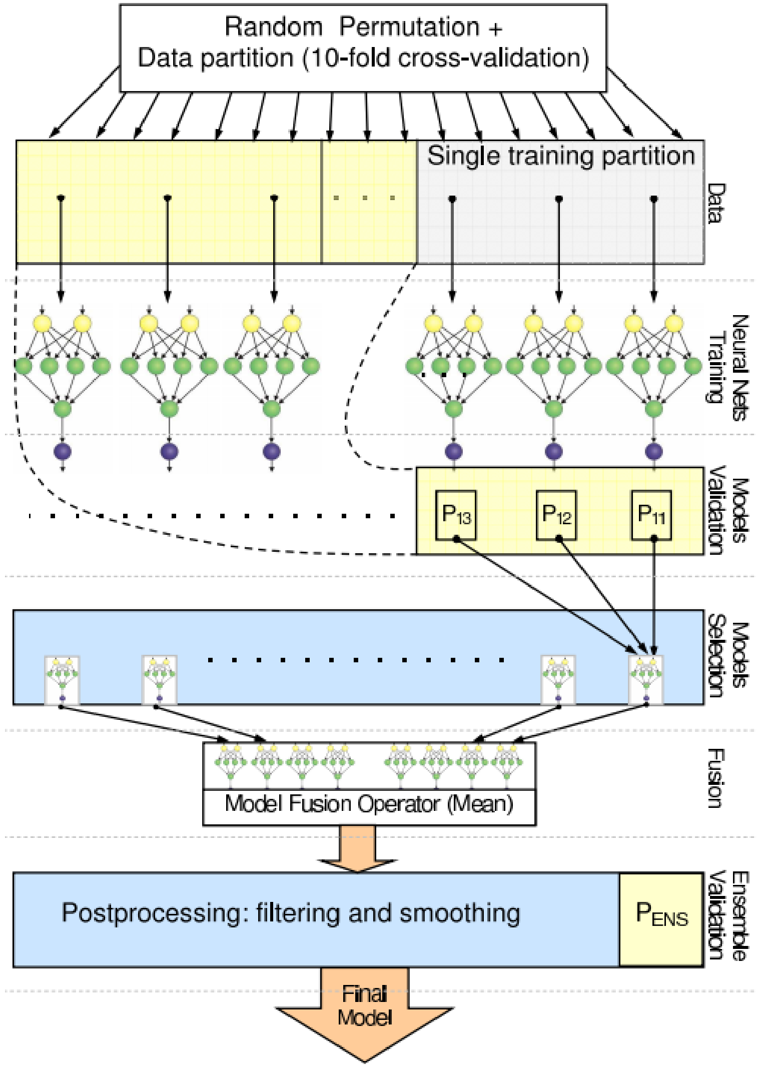
- OUR SOLUTION (The best nature inspired concept):
  - Recurrent neural network trained and adapted via dynamic particle swarm optimization




STEP 1
STEP 2
STEP 3
STEP 4
Random initialization of complete swarm
Random initialization of half of swarm
Random initialization of half of swarm
Random initialization of half of swarm
Training
Adapting
Adapting
Adapting
Data from Step1
Data from Step2
Data from Step3
Data from Step4

Prediction – chemical reactor activity

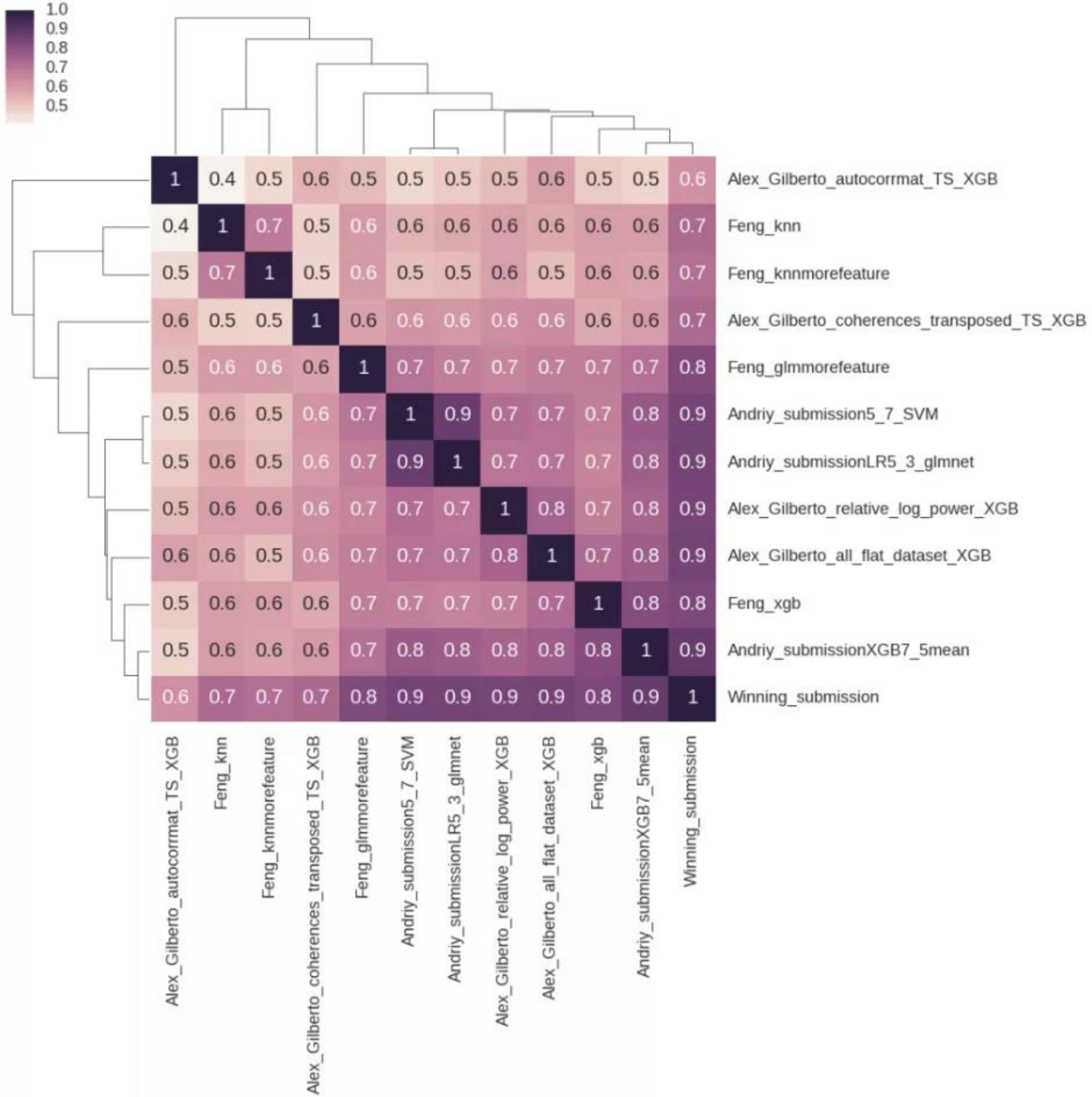
- Most accurate solution:
  - Dymitr Ruta/Bogdan Gabrys
  - Bournemouth university
  - Ensemble of BP neural networks


Random Permutation +
Data partition (10-fold cross-validation)
Single training partition
Data
Neural Nets Training
Models Validation
P13
P12
P11
Models Selection
Fusion
Model Fusion Operator (Mean)
Postprocessing: filtering and smoothing
PENS
Ensemble Validation
Final Model

Classification - Seizure Prediction Competition

- From an interview with winner team:
  - bunch of different classifiers (XGB, SVM, KNN, LR).
  - simple solutions with minimal parameter tuning.
  - diversity in the ensemble is the key to robustness
  - many simple and relatively low performing models rather than trying to hyper-optimize our best performing models (and overfit in the process).
  - simple feature sets works very well in this dataset.
  - when CV is unreliable, don’t panic, simple things and basic ensembling (and teaming) provide a very stable solution.

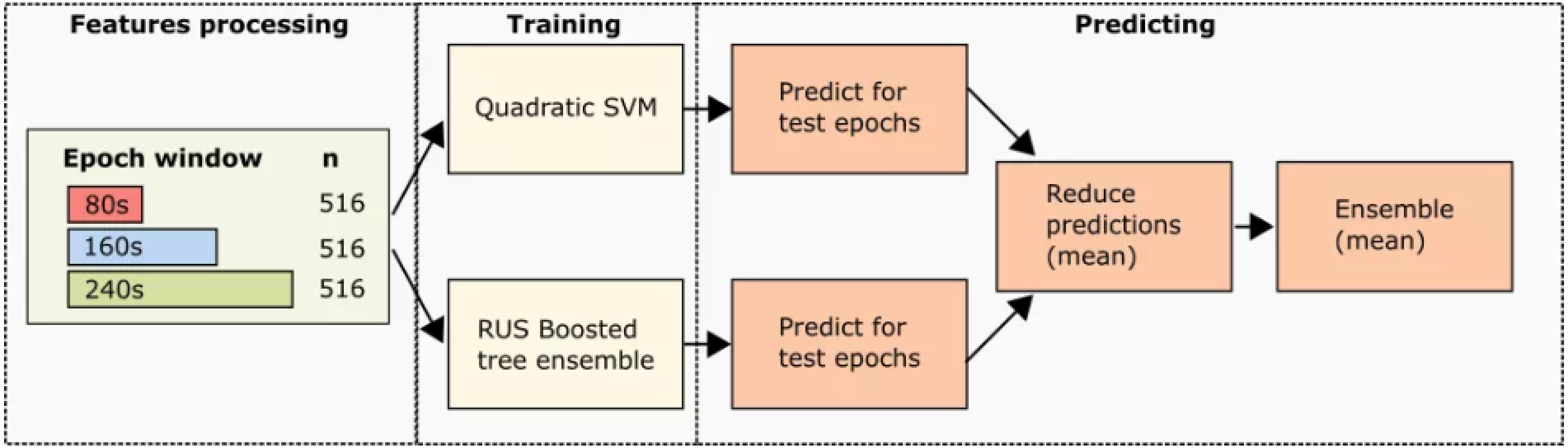
CORRELATION MAP BETWEEN EACH OF THE 11 INDIVIDUAL MODELS AND THE WINNING SOLUTION. THE OVERALL LOW CORRELATION SHOW A STRONG DIVERSITY IN THE MODELS PREDICTIONS.



	Alex_Gilberto_autocorrmat_TS_XGB	Feng_knn	Feng_knnmorefeature	Alex_Gilberto_coherences_transposed_TS_XGB	Feng_glmmorefeature	Andriy_submission5_7_SVM	Andriy_submissionLR5_3_glmnet	Alex_Gilberto_relative_log_power_XGB	Alex_Gilberto_all_flat_dataset_XGB	Feng_xgb	Andriy_submissionXGB7_5mean	Winning_submission
Alex_Gilberto_autocorrmat_TS_XGB	1	0.4	0.5	0.6	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.6
Feng_knn	0.4	1	0.7	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7
Feng_knnmorefeature	0.5	0.7	1	0.5	0.6	0.5	0.5	0.6	0.5	0.6	0.6	0.7
Alex_Gilberto_coherences_transposed_TS_XGB	0.6	0.5	0.5	1	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7
Feng_glmmorefeature	0.5	0.6	0.6	0.6	1	0.7	0.7	0.7	0.7	0.7	0.7	0.8
Andriy_submission5_7_SVM	0.5	0.6	0.5	0.6	0.7	1	0.9	0.7	0.7	0.7	0.8	0.9
Andriy_submissionLR5_3_glmnet	0.5	0.6	0.5	0.6	0.7	0.9	1	0.7	0.7	0.7	0.8	0.9
Alex_Gilberto_relative_log_power_XGB	0.5	0.6	0.6	0.6	0.7	0.7	0.7	1	0.8	0.7	0.8	0.9
Alex_Gilberto_all_flat_dataset_XGB	0.6	0.6	0.5	0.6	0.7	0.7	0.7	0.8	1	0.7	0.8	0.9
Feng_xgb	0.5	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	1	0.8	0.8
Andriy_submissionXGB7_5mean	0.5	0.6	0.6	0.6	0.7	0.8	0.8	0.8	0.8	0.8	1	0.9
Winning_submission	0.6	0.7	0.7	0.7	0.8	0.9	0.9	0.9	0.9	0.8	0.9	1

Classification - Seizure Prediction Competition

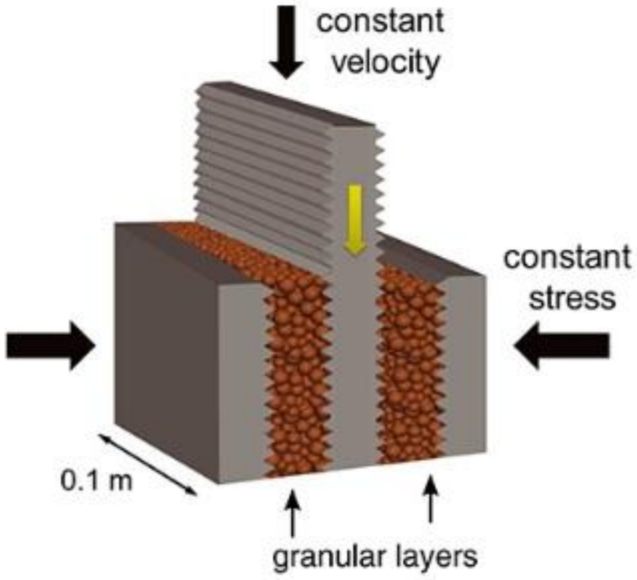
- From and interview with Gareth Jones, 3rd Place Winner:
  - the public leaderboard used only 30% of the test data, so overfitting was a huge risk (the final top ten for this competition had a net position gain of more than 100)
  - ensemble of a quadratic SVM and an RUS boosted tree ensemble with 100 learners

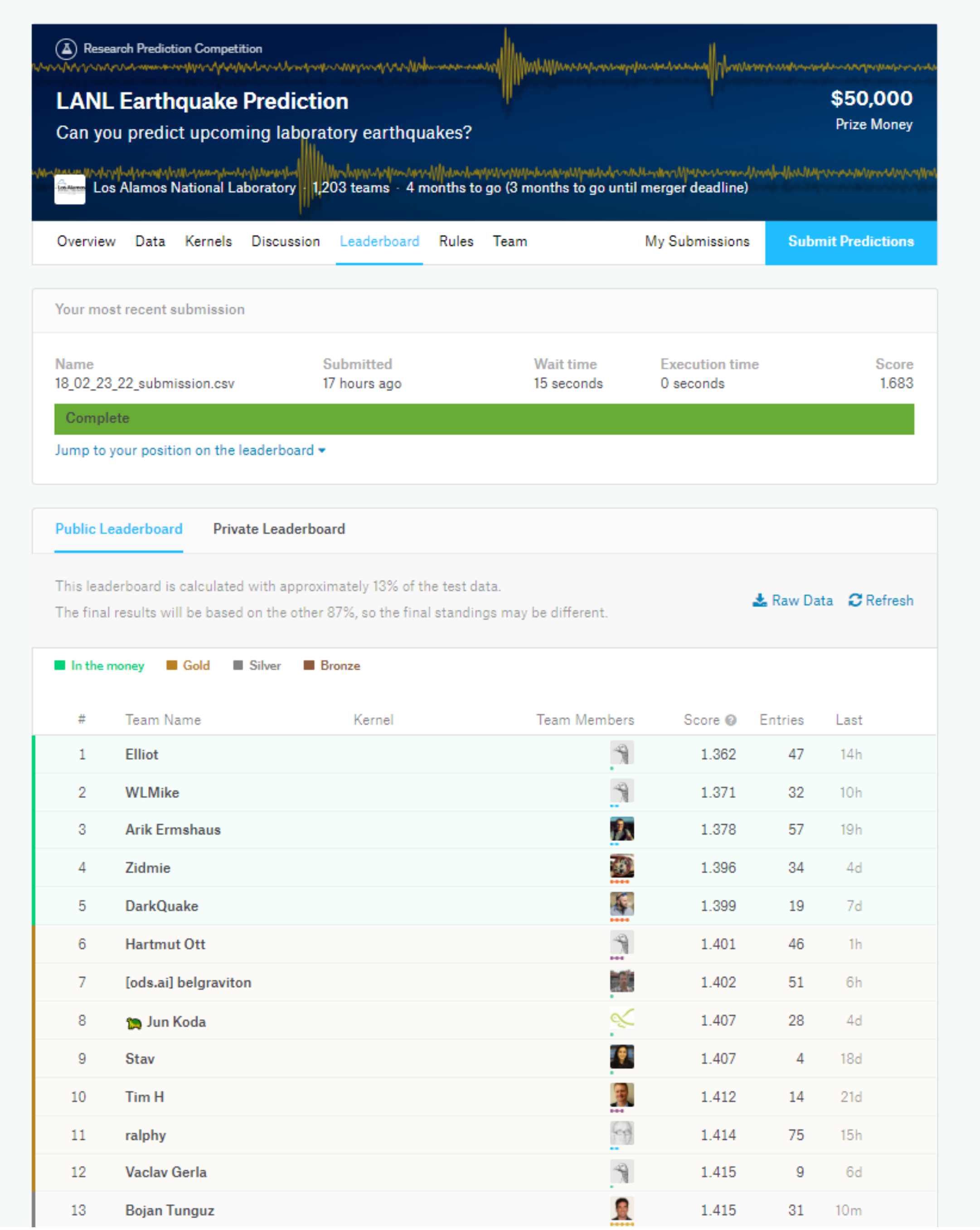


Source: http://blog.kaggle.com/2017/01/10/seizure-prediction-competition-3rd-place-winners-interview-gareth-jones/

Regression – earthquake prediction

- Currently active competition
- Laboratory (simulated) earthquakes only
- https://www.kaggle.com/c/LANL-Earthquake-Prediction




Research Prediction Competition

LANL Earthquake Prediction

Can you predict upcoming laboratory earthquakes?

$50,000 Prize Money

Los Alamos National Laboratory · 1,203 teams · 4 months to go (3 months to go until merger deadline)

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
18_02_23_22_submission.csv	17 hours ago	15 seconds	0 seconds	1.683

Complete

Jump to your position on the leaderboard ▾

This leaderboard is calculated with approximately 13% of the test data.
The final results will be based on the other 87%, so the final standings may be different.

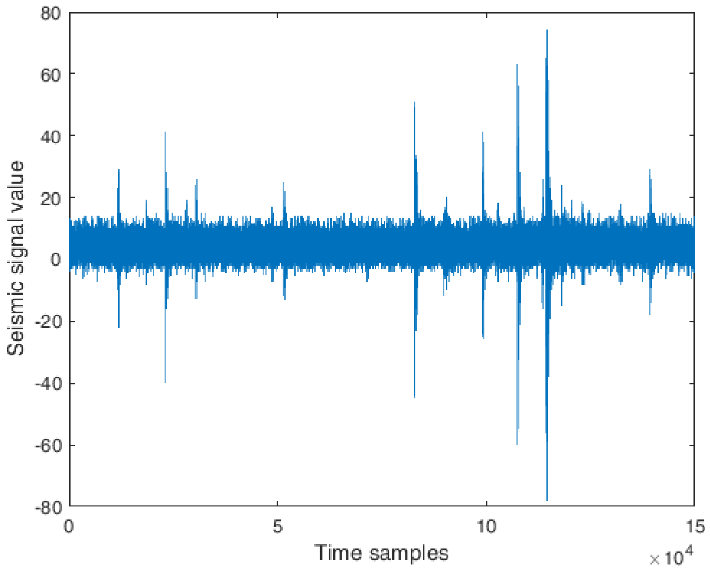
■ In the money ■ Gold ■ Silver ■ Bronze

#	Team Name	Kernel	Team Members	Score	Entries	Last
1	Elliot			1.362	47	14h
2	WLMike			1.371	32	10h
3	Arik Ermshaus			1.378	57	19h
4	Zidmie			1.396	34	4d
5	DarkQuake			1.399	19	7d
6	Hartmut Ott			1.401	46	1h
7	[ods.ai] belgraviton			1.402	51	6h
8	Jun Koda			1.407	28	4d
9	Stav			1.407	4	18d
10	Tim H			1.412	14	21d
11	ralphy			1.414	75	15h
12	Vaclav Gerla			1.415	9	6d
13	Bojan Tunguz			1.415	31	10m

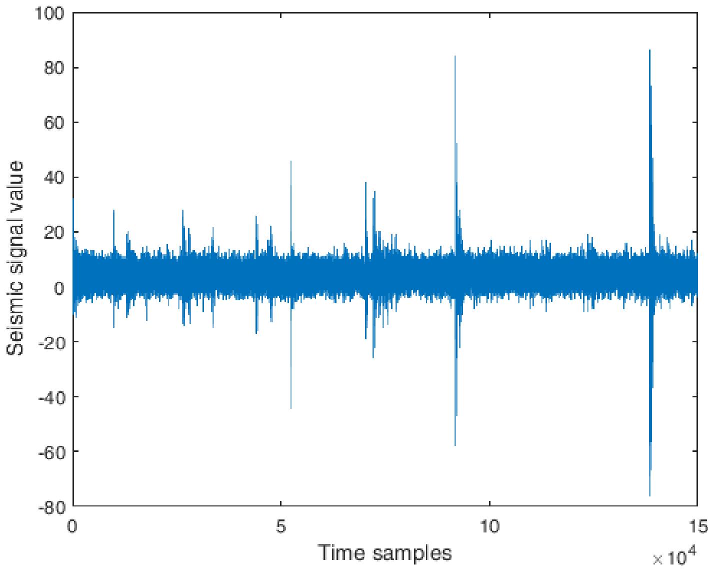
Regression – earthquake prediction

- The goal
  - to use seismic signals to predict the timing of laboratory earthquakes
  - data comes from a well-known experimental set-up used to study earthquake physics
  - input seismic/acoustic signal is used to predict the time remaining before the next laboratory earthquake

8 seconds to quake

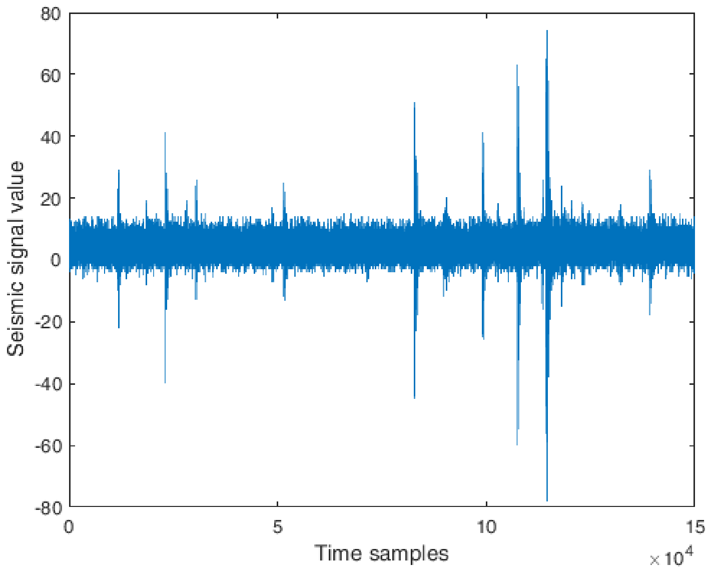


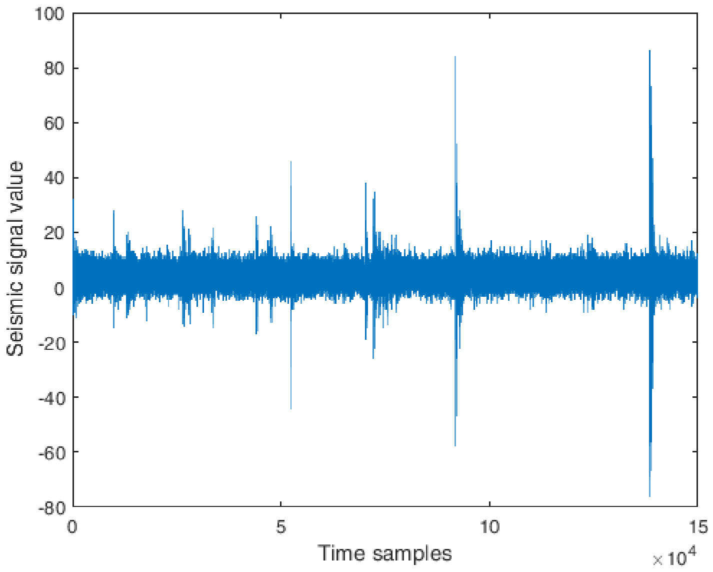
0.06 seconds to quake



Regression – earthquake prediction

- Current solutions
  - Secret☺



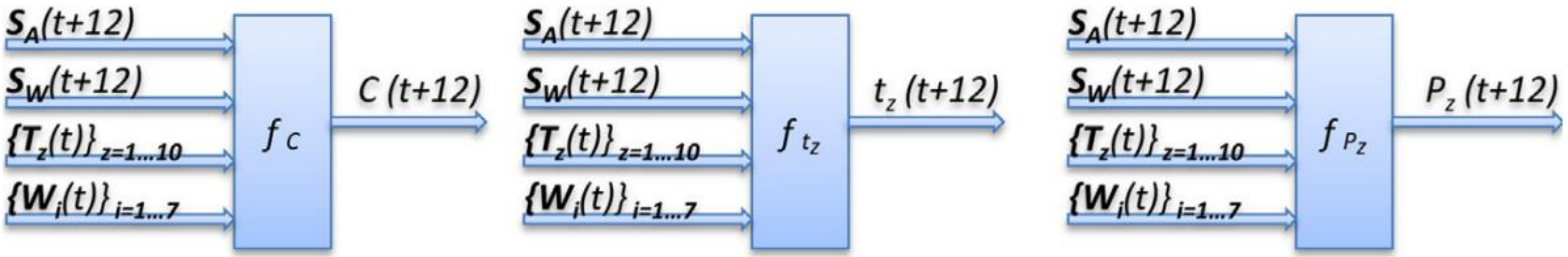


Lessons learned from competitions

- Ensemble methods that combine multiple models mostly win competitions
- Simpler methods are often robust and win competitions with higher overfitting risk
- Data preprocessing, feature extraction and feature selection are critical
- A good performance estimate/validation methodology is critical

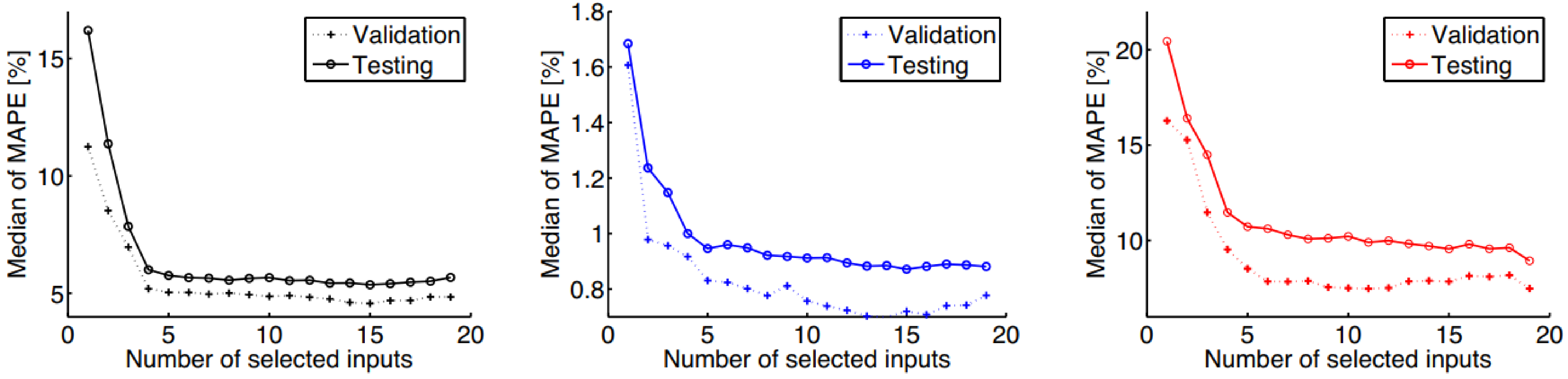
Other related applications

Smart building heating - Anomaly detection



$S_A(t+12)$
$S_W(t+12)$
$\{T_z(t)\}_{z=1\ldots10}$
$\{W_i(t)\}_{i=1\ldots7}$
f_C
$C(t+12)$
$S_A(t+12)$
$S_W(t+12)$
$\{T_z(t)\}_{z=1\ldots10}$
$\{W_i(t)\}_{i=1\ldots7}$
f_{t_z}
$t_z(t+12)$
$S_A(t+12)$
$S_W(t+12)$
$\{T_z(t)\}_{z=1\ldots10}$
$\{W_i(t)\}_{i=1\ldots7}$
f_{P_z}
$P_z(t+12)$

Consumption model *Temperature models* *Discomfort models*



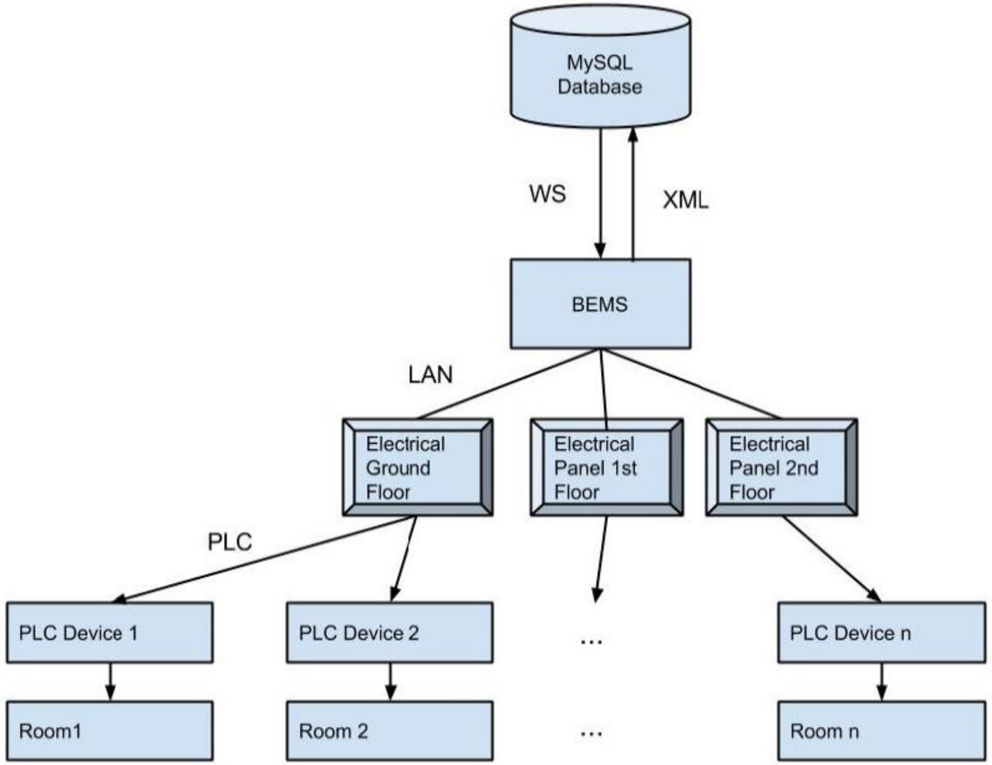
(a) Consumption (b) Temperature (c) Discomfort

Smart building heating - Anomaly detection

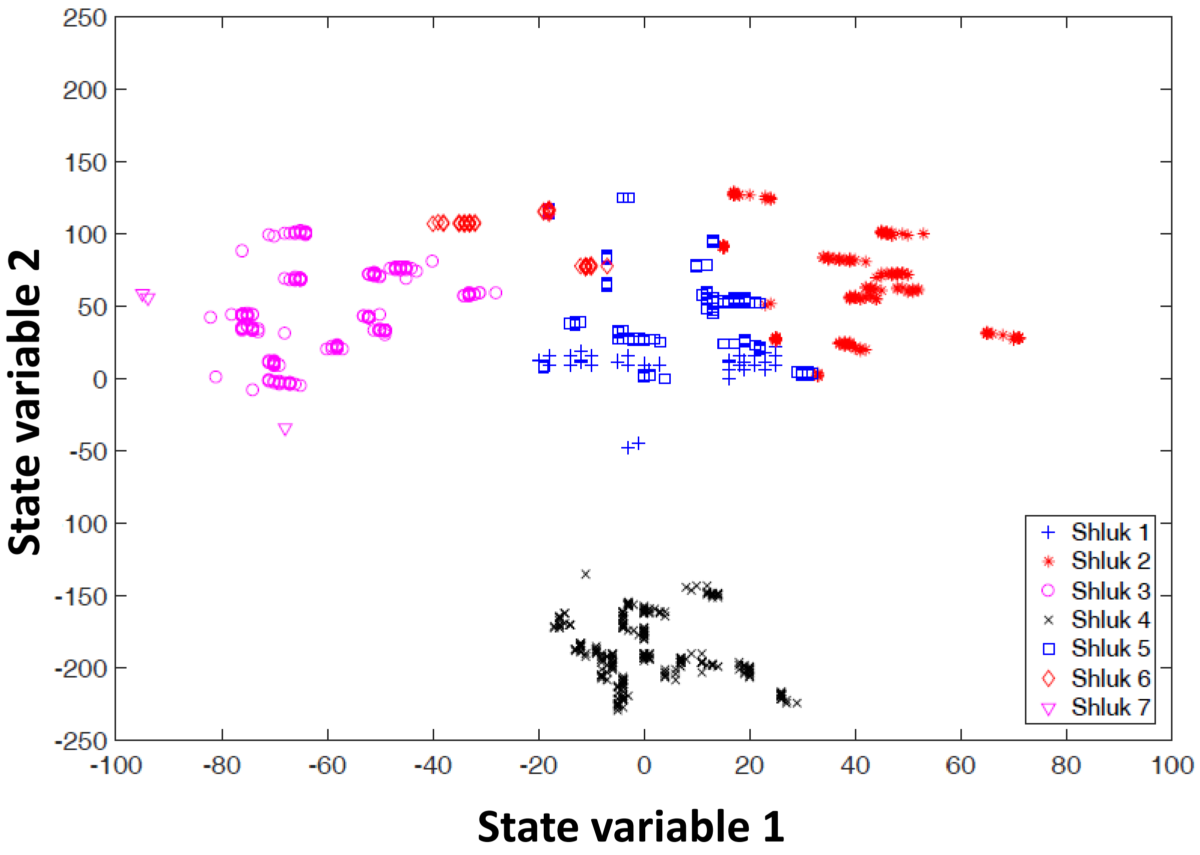
- The importance of inputs selection in HVAC modeling is pointed out and demonstrated
- It is observed that the early stopping mechanism is crucial especially but not only for small training data, because it reliably overcomes overfitting problems.
- Macas, M., Moretti, F., Fonti, A., Giantomassi, A., Comodi, G., Annunziato, M., ... & Capra, A. (2016). The role of data sample size and dimensionality in neural network based forecasting of building heating related variables. *Energy and Buildings*, *111*, 299-310.

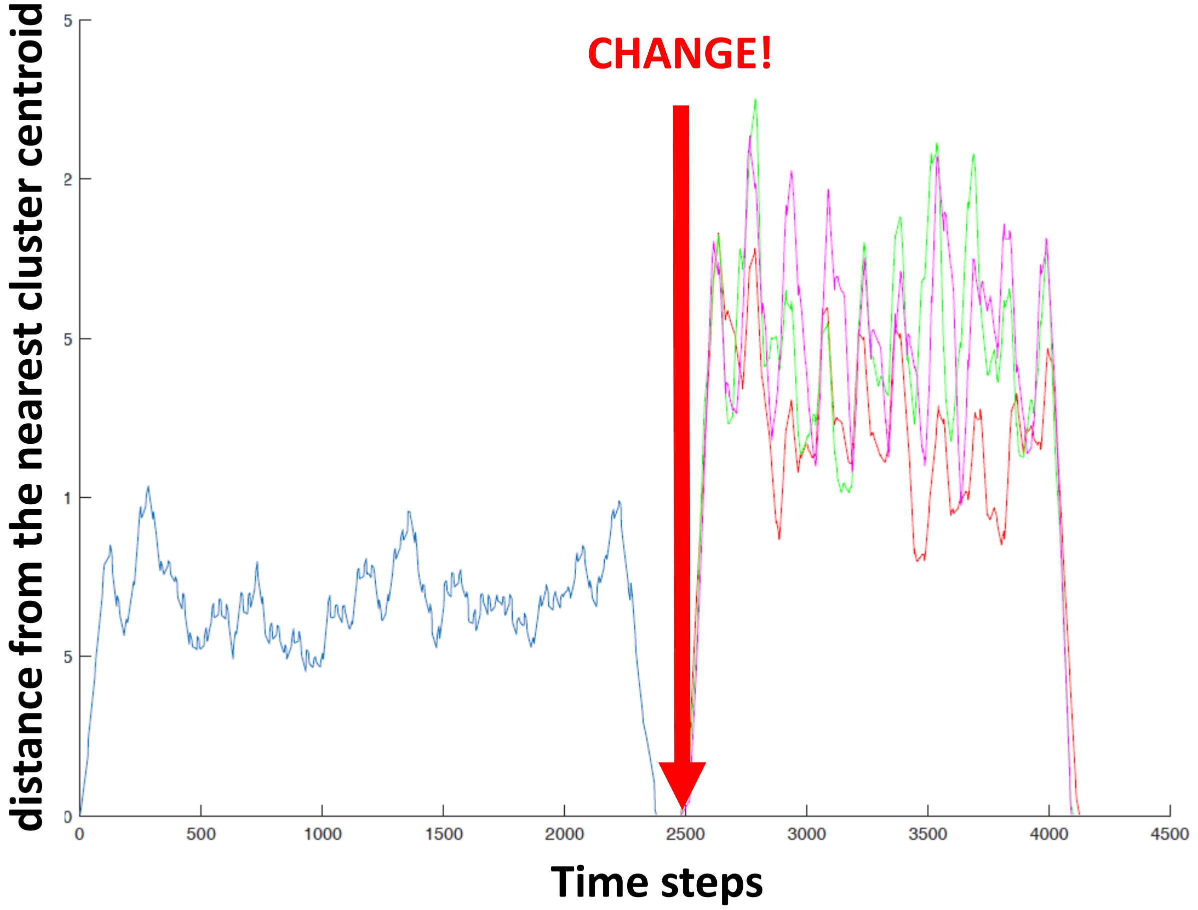
Anomaly detection – smart building heating

- Office building located at ENEA Research Centre (Rome, Italy)



Anomaly detection – smart building heating





CHANGE!

Anomaly detection – smart building heating

- Other approaches: Peak detection and fuzzy rules

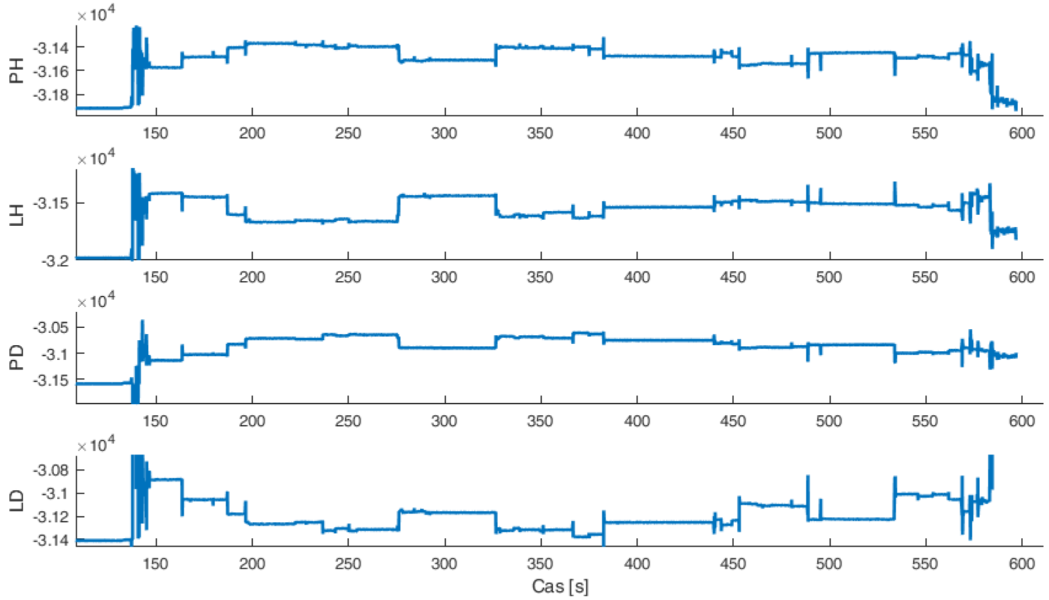
- Lauro, F., Moretti, F., Capozzoli, A., Khan, I., Pizzuti, S., Macas, M., & Panzieri, S. (2014). Building fan coil electric consumption analysis with fuzzy approaches for fault detection and diagnosis. *Energy Procedia*, *62*, 411-420.

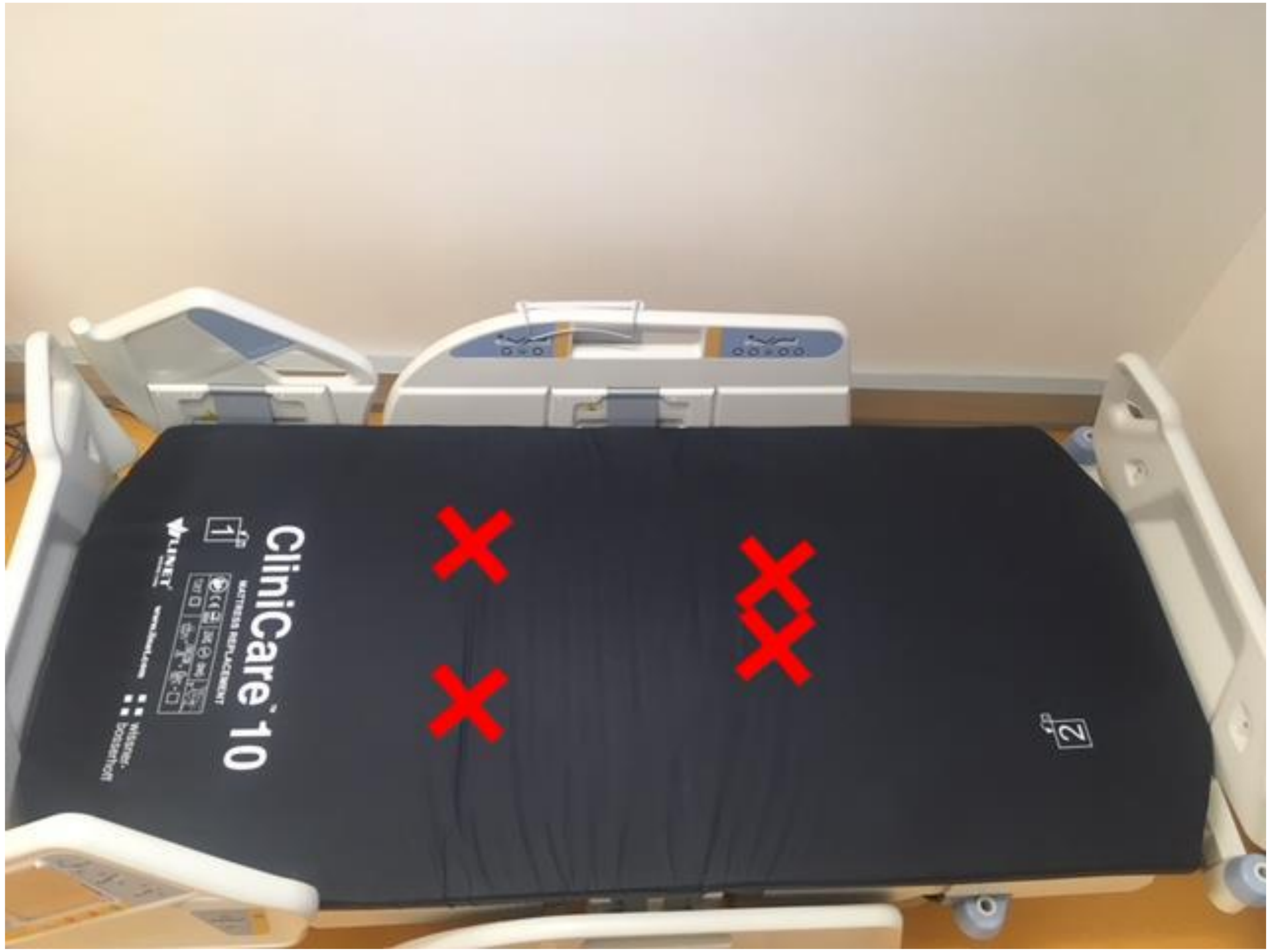
Change detection – smart bed

- Collaboration with LINET company
- Goal:
  - To detect a significant change

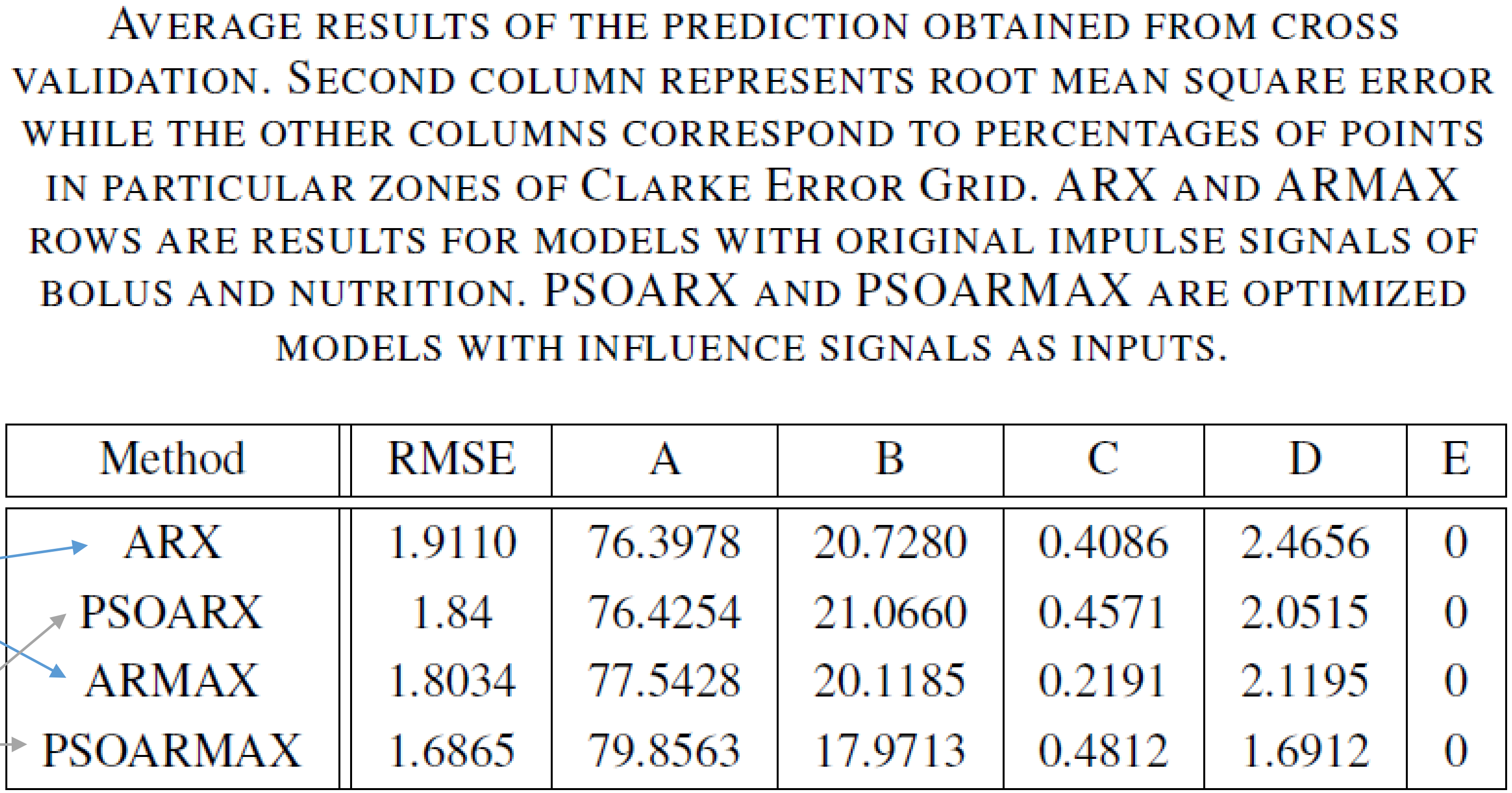
Change detection – smart bed

- Inputs:
  - Signals from four strain gauges





Glycaemia forecasting



Average results of the prediction obtained from cross validation. Second column represents root mean square error while the other columns correspond to percentages of points in particular zones of Clarke Error Grid. ARX and ARMAX rows are results for models with original impulse signals of bolus and nutrition. PSOARX and PSOARMAX are optimized models with influence signals as inputs.

Method	RMSE	A	B	C	D	E
ARX	1.9110	76.3978	20.7280	0.4086	2.4656	0
PSOARX	1.84	76.4254	21.0660	0.4571	2.0515	0
ARMAX	1.8034	77.5428	20.1185	0.2191	2.1195	0
PSOARMAX	1.6865	79.8563	17.9713	0.4812	1.6912	0

Scenario 1 – raw data as inputs

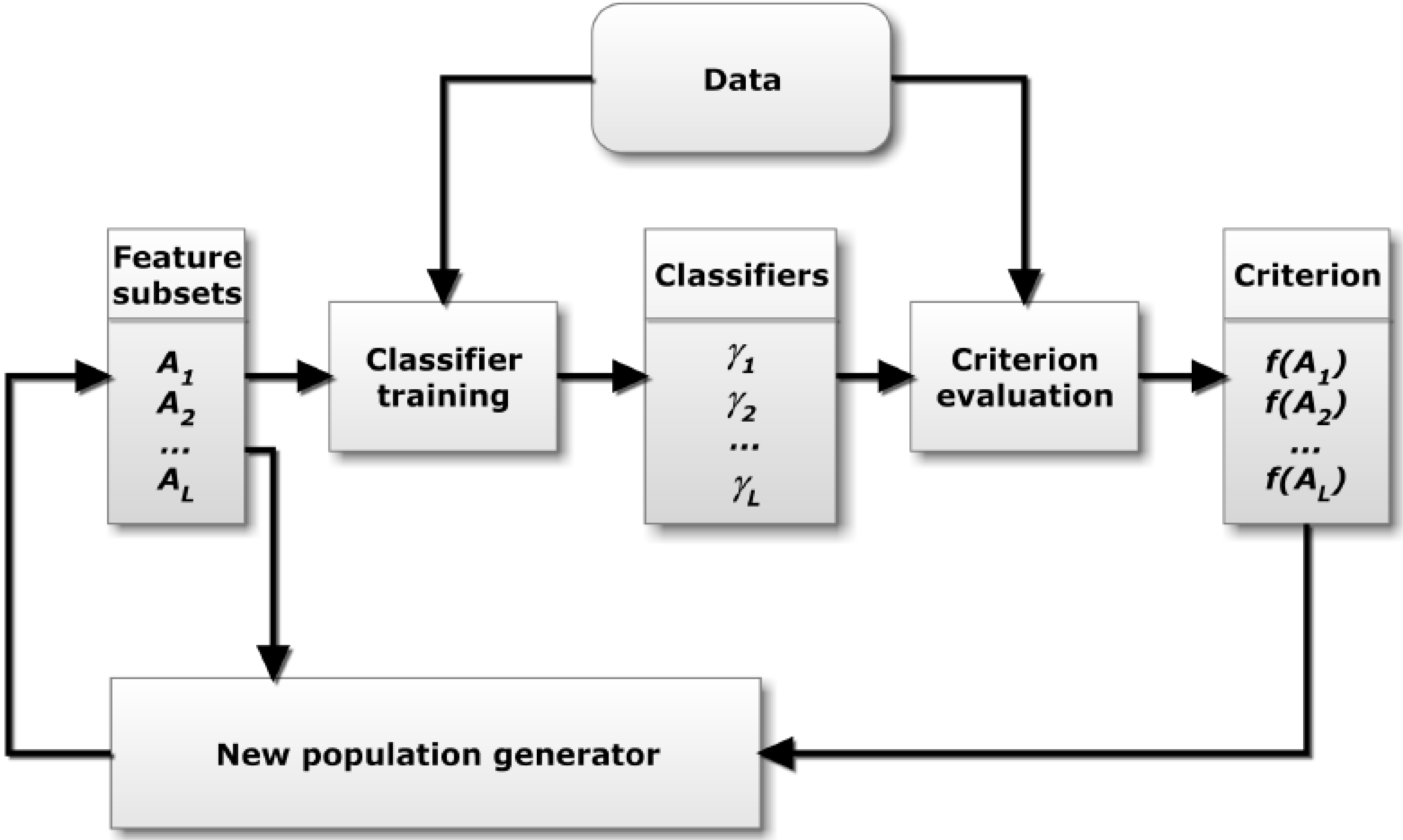
Scenario 2 – PSO optimized models of effects

Conclusions

- Our experience shows that although a domain knowledge is important, ML can be quickly and succesfully applied
- Currently, ML community is becoming more and more useful in most application areas
- ML competitions provide very important knowledge about ML state-of-the-art and are more important than journal or conference papers biased by publish-or-perish pressure
- ML model type and its learning is typically not the most important part of a data modelling process.
- Model ensembles that combine multiple machine learning models are the real STATE-OF-THE-ART.

Thank You!

Wrapper feature selection



Main problem

- Big sample size
  - the population based heuristic search is time consuming
- Small sample size
  - The error estimates have high variance
  - The feature selection criterion is inaccurate
  - High feature selection bias
  - We minimize something, which is different from the true error
- Solution:
  - reduce the variance of the error estimate

Complete error estimates for nearest neighbor classifier

- Complete error estimates
  - error estimates averaged over all random partitions into the training and testing set
  - **1-nearest neighbor classifier** was focused
  - Complete cross-validation (Mullin, 2000) for 1NN was **applied**
  - Complete bootstrap (Macas, 2012) for 1NN was **introduced**