DSC 5.0 Tutorial:
End-to-End ML using h2o in R

Belgrade, November 2019.

About me

- Owner & Consultant @ Logikka
- Senior Data Scientist @ NIS Gazprom Neft
- Board Member @ Data Science Serbia
- Proud ETF alumni

Get tutorial materials:

https://bit.ly/dsc50_h2o_tutorial

WHO ARE YOU?

What is h2o?

H_2O

In-memory, distributed machine learning algorithms

Spark + H_2O

SPARKLING WATER

Integration with Apache Spark

H_2O4GPU

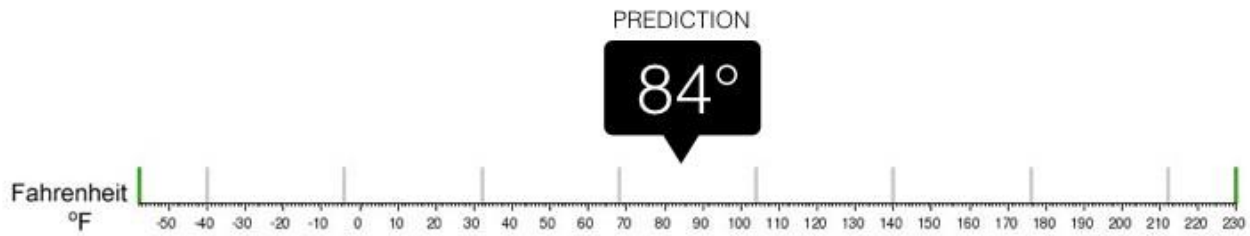
Machine Learning on GPUs

100% Open Source

What is Machine Learning?

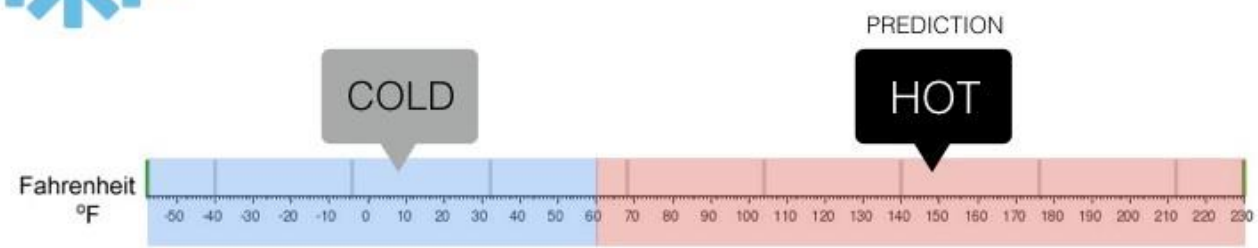
Regression

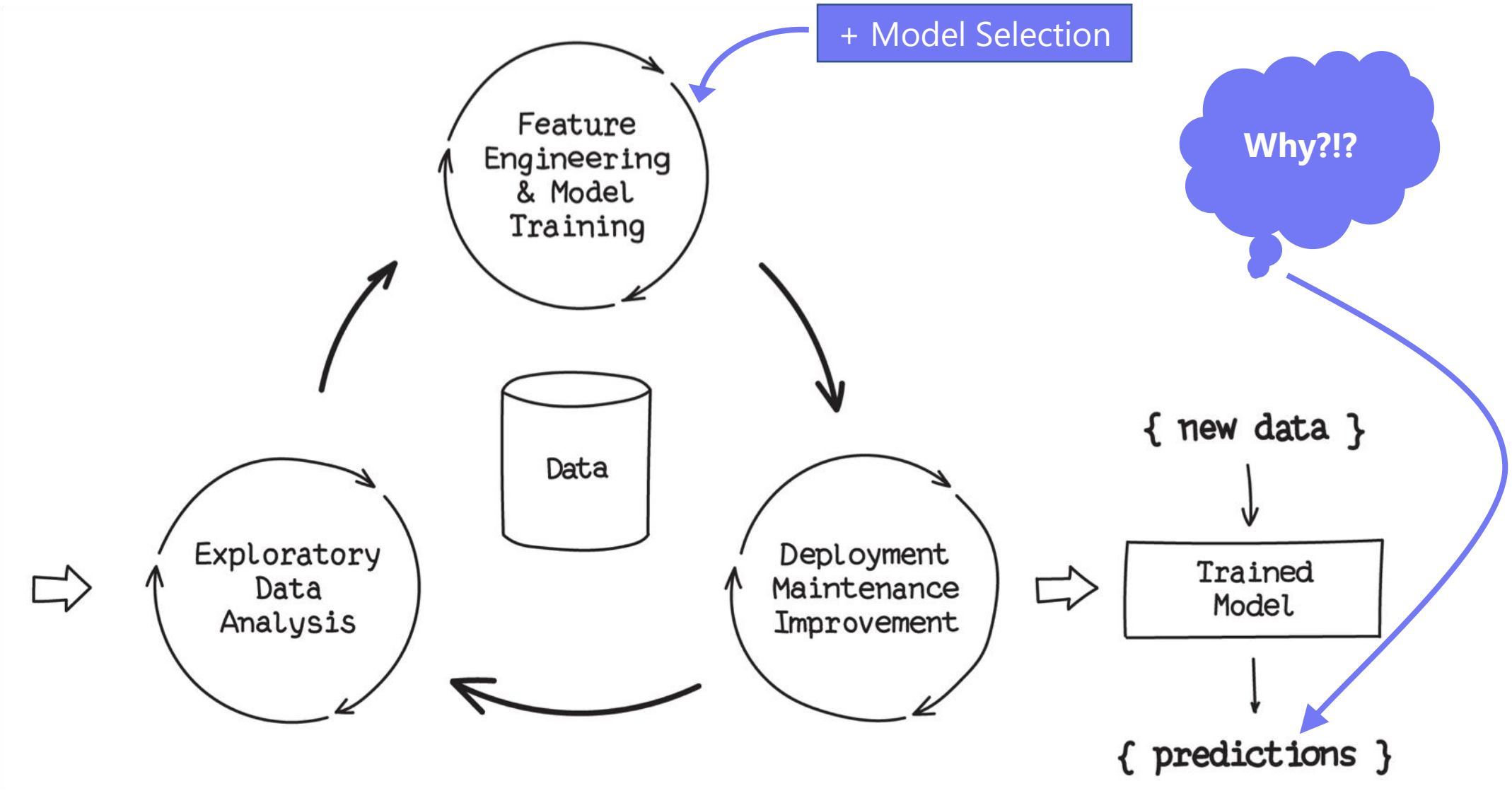
What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?

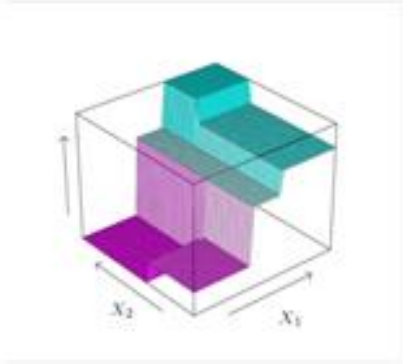



+ Model Selection
Why?!?
Feature Engineering & Model Training
Data
Exploratory Data Analysis
Deployment Maintenance Improvement
{ new data }
Trained Model
{ predictions }

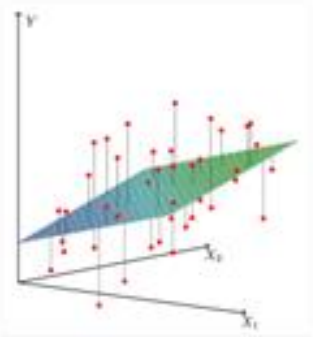
Supported ML Algorithms

DRF
XRT

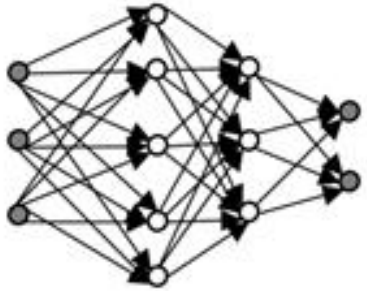
GBM
XGBoost



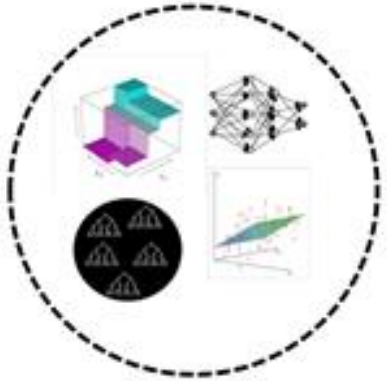
GLM



DNN



Stacked
Ensemble




9.1
9.0
9.0
7.2

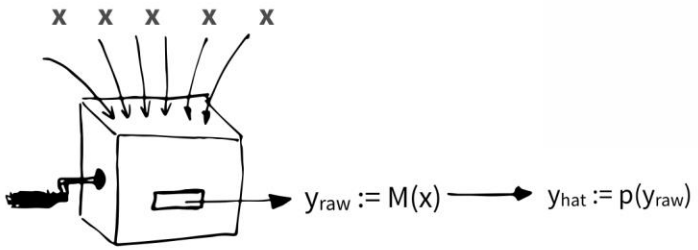
DALEX - Descriptive mAchine Learning EXplanations

DALEX explains black-box models. It's a methodology for better diagnostic of any black-box model.

This approach increases understanding of a model, increases trust in model predictions and allows to further improve the model. It also allows to compare two or more models in the scale space

Notation:

- **(x, y)** - pair of input and output data points. x may be anything (data.frame, factors, numbers, text, image), while here we assume that y is numerical or can be transformed to the numerical variable ($x \in X$; $y \in R$).
- **M** - a black box model, $M: X \rightarrow R$. Its output will be denoted as $y_{raw} = M(x)$
- **p** - a link function, transforms raw model output to the same space as y. Useful for classification, while for regression its usually the identity. $p: R \rightarrow R$. Its output will be denoted as $y_{hat} = p(y_{raw})$



model M raw output final output

explain(model; data; y; predict_function; trans)

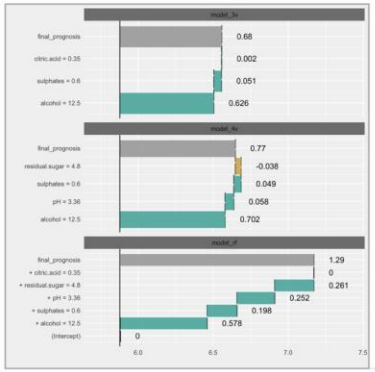
The *explain*() function creates a wrapper over a black-box model. This wrapper contains all necessary components for further processing.

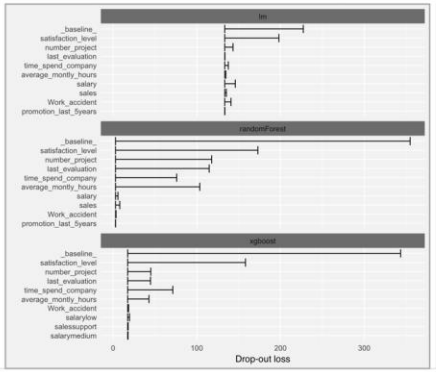
prediction_breakdown(explainer, x) variable_importance(explainer) variable_response(explainer, variable)

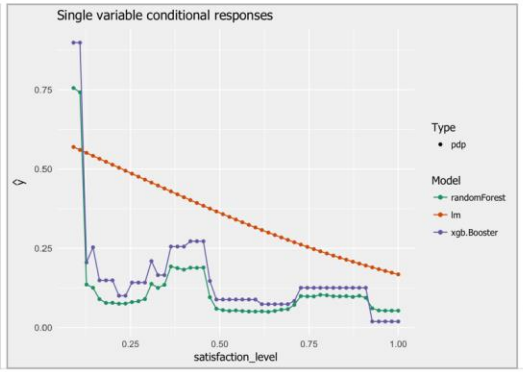
Prediction explainers shows features that drive model response for a selected observation

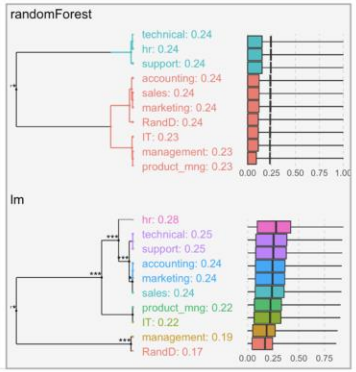
Variable importance explainers shows the drop in the model loss after permutations of a selected variable.

Single variable explainers show conditional relation between model output and a single variable.









Thank you!

branko@logikka.ai
www.linkedin.com/in/kovacbranko
www.logikka.ai

LOGIKKA