R/Pharma 2021

Industrialized Machine Learning and Explainable AI for Late Phase Trials

Biomarker & Data Insights

Data Science & Analytics

Karl Köchert	bayer/bmdi
Adam Skubala	bayer/bmdi
Antigoni Elefsinioti	bayer/bmdi
Eliana Garcia Cossio	bayer/bmdi
Frauke Hennig	chrestos/bsds
Lidia Sacchetto	bayer/cseu
Lukas Lasota	clinstat
Maike Ahrens	chrestos/bsds
Sebastian Voss	chrestos/bsds
Yuan Hong	bayer/bmdi

BMDI MLAI pipeline

WHY

Why do we need the MLAI pipeline?

WHAT

What is the MLAI pipeline?

HOW

How does it work?
How was the pipeline built?

Today's aim:

Briefly touch the Why, What and the Hows.

Why do we need the MLAI pipeline?

Motivation and added value of MLAI in late phase interventional trials

Classical approaches are "forcing" a hypothesis on the data and are limited to detection of strong, low-complexity signals using only a small amount of the variables (measured data) at one time.

By deriving the best hypothesis given the data, Machine Learning (ML) is currently the best available methodology to create holistic mathematical models of complex (biological) systems using all available data and variables.

In that sense, ML complements classical approaches to ensure prudent and as thorough as technically possible insights generation.

Explainable AI needed – focus on understanding data-outcome relations, not prediction!

Why do we need the MLAI pipeline?

Motivation and added value of using a MLAI pipeline

Standardized
to enable transparency as well as efficiency and the reusability across many projects.

Reproducible
at every step to guarantee consistency, robustness and trust.

Validated
to follow the general principles of software validation (EMA / FDA).

What is the MLAI pipeline?

We provide a standardized, reproducible and validated MLAI pipeline to efficiently compute and report holistic MLAI models for clinical endpoints (classification, regression, time-to-event) based on all measured data in a clinical trial.

The pipeline's purpose is to detect clinically relevant, highly complex safety / efficacy signals that are not identifiable with classical approaches.

Find answers to questions such as

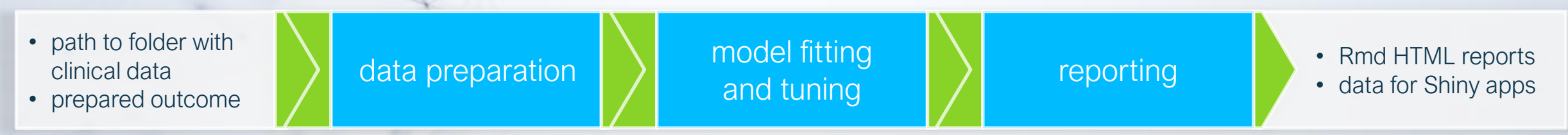
- How good is the measured data at explaining a clinical outcome / an endpoint of interest?
- What are potential drivers / features for the occurrence of an outcome?
- Do we find baseline characteristics that interact with treatment regarding the effect on the outcome?
- Can we identify potential directionality / relationship between feature values and the outcome?

The pipeline is already being used in various Bayer studies, MLAI topline close to study topline results!

How does it work? How was the pipeline built?

General overview

A bunch of **R functions & Rmd templates**, summarized into **R packages**, summarized into the **meta package** *bmdiMARTINI*



Flexible modules
to address questions regarding the association between an outcome of interest and the collected data.

Expandable modules
for future needs.

Validated modules
so that it can be used in the clinical setting.

How does it work? How was the pipeline built?

Module concepts

- path to folder with clinical data
- prepared outcome

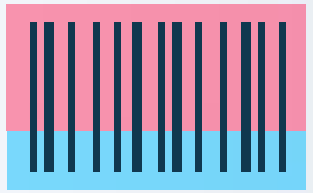
data preparation → model fitting and tuning → reporting → • Rmd HTML reports • data for Shiny apps

clinical data base

study data collected in several SAS-files (ADaM | CDISC)

ML readable format (tidy)

prepared ML data



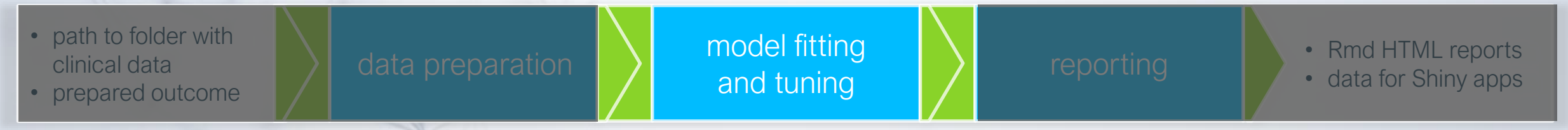
train / test split

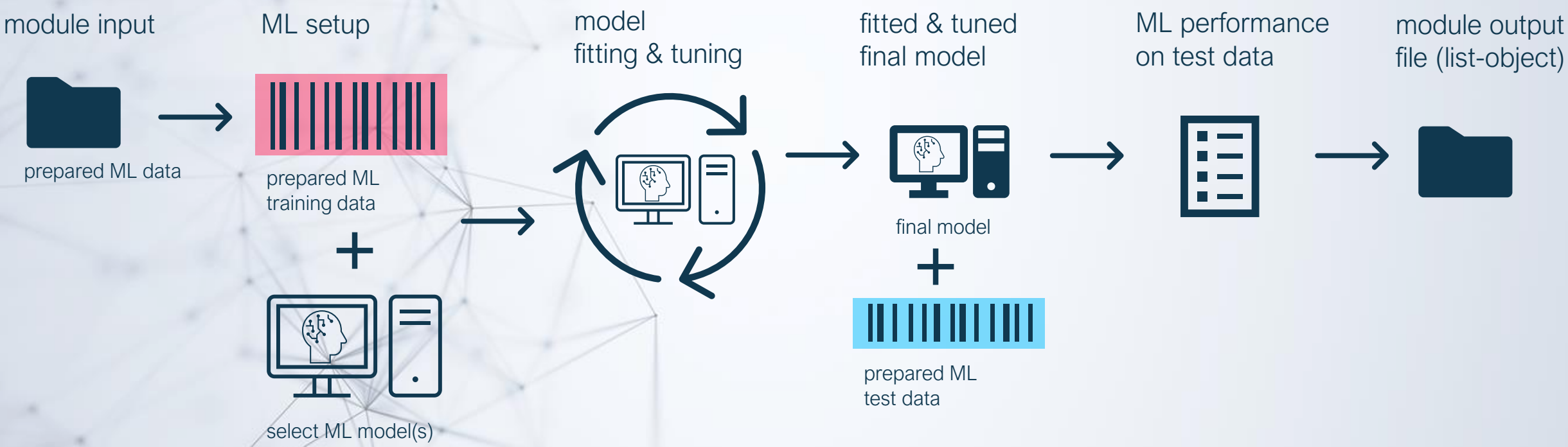
module output file (list-object)

```
1  path_to_ads %>%
2    ads_spec() %>%
3    build() %>%
4    prepare_ml()
```

How does it work? How was the pipeline built?

Module concepts





BMDI

How does it work? How was the pipeline built?

Module concepts

- path to folder with clinical data
- prepared outcome

data preparation

model fitting and tuning

reporting

- Rmd HTML reports
- data for Shiny apps

module input

final model

prepared ML data

run sub-modules

data summaries

RF explainer

variable importance

2-way interaction

partial dependence / accumulated-local profiles

...

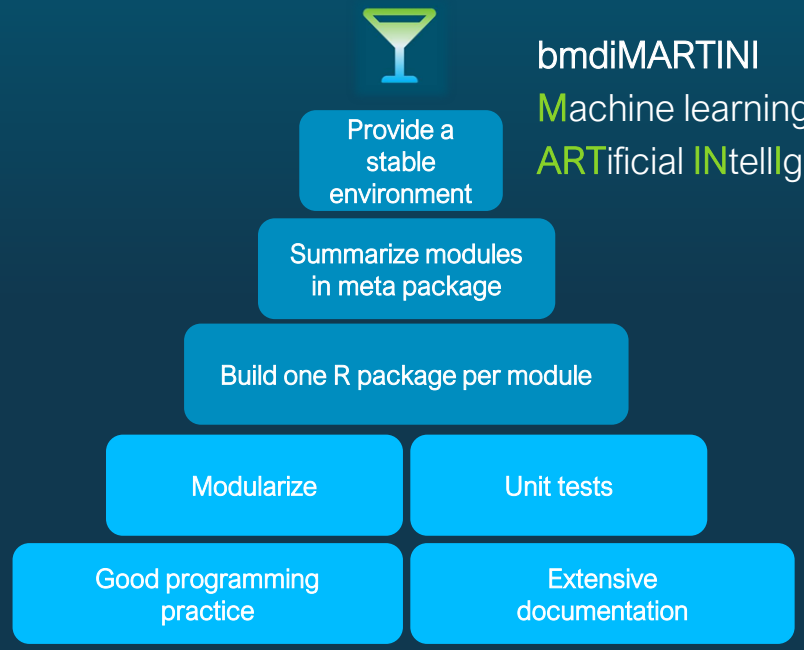
modules output / input

use Rmd templates

Short demo

Thank you!

R/Pharma 2021



bmdiMARTINI

Machine learning &
ARTificial INtelligence