Supplementary Information - Analysis Code

Table of Contents

suppressPackageStartupMessages(library(tidyverse))  
library(targets)  
library(tarchetypes)  
library(visNetwork)

knitr::opts\_knit$set(root.dir = "../../../", tar\_interactive = FALSE, collapse = TRUE, comment = "#>")  
tar\_unscript()

# 1 Disclaimer

This document only contains the high-level workflow. For a full documentation of the entire code base please visit the public code repository on github and the accompanying website for a readable report on the data structure and code results.

* <https://github.com/openpharma/quality_risk_assesment_clinical_trials>
* <https://openpharma.github.io/quality_risk_assesment_clinical_trials/index.html>

# 2 {targets} Workflow

## 2.1 Globals

options(tidyverse.quiet = TRUE)  
tar\_option\_set(  
 packages = c(  
 "gt",  
 "yardstick",  
 "broom",  
 "cowplot",  
 "glue",  
 "tidyverse",  
 "arrow",  
 "tarchetypes"  
 )  
 )  
purrr::map(dir("./src/R", full.names = TRUE), source)

## Established \_targets.R and \_targets\_r/globals/globals.R.

## 2.2 Targets

### 2.2.1 Config

tar\_target(  
 config,  
 list(  
 # columns that are not features  
 id\_vars = c(  
 'category\_id',  
 'modelling\_category',  
 'classification',  
 'activity\_id\_new',  
 'start\_date',  
 'index',  
 'date\_first\_study\_activity',  
 'date\_last\_study\_activity',  
 'protocol',  
 'file\_name',  
 'source\_row\_index',  
 'audit\_or\_inspection',  
 'site\_platinum\_id',  
 'pi\_platinum\_id',  
 'site\_num',  
 'country',  
 'site\_activation\_date',  
 'site\_closed\_date',  
 'site\_closed\_date\_corr',  
 'study\_start\_date',  
 'study\_end\_date'  
 ),  
 max\_year = 2020  
 )  
)

## Established \_targets.R and \_targets\_r/targets/config.R.

### 2.2.2 Data

#### 2.2.2.1 Files

list(  
 tar\_target(file\_mm, "data/in/modelling\_matrix.feather", format = "file"),  
 tar\_target(file\_mm\_bin, "data/in/lasso\_prep.feather", format = "file"),  
 tar\_target(file\_form, "data/in/glm\_coefs.feather", format = "file"),  
 tar\_target(file\_cv, "data/in/indeces\_annual\_splits.feather", format = "file"),  
 tar\_target(file\_cat\_lookup, "data/in/category\_lookup.feather", format = "file"),  
 tar\_target(file\_feat\_lookup, "data/in/feature\_lookup.csv", format = "file")  
)

## Established \_targets.R and \_targets\_r/targets/files.R.

#### 2.2.2.2 Load

list(  
 tar\_target(df\_mm, qract\_read\_and\_anonymize(file\_mm, config$id\_vars, arrow::read\_feather)),  
 tar\_target(df\_mm\_bin, qract\_read\_and\_anonymize(file\_mm\_bin, config$id\_vars, arrow::read\_feather)),  
 tar\_target(df\_form, qract\_read\_and\_anonymize(file\_form, config$id\_vars, arrow::read\_feather)),  
 tar\_target(df\_cv, qract\_read\_and\_anonymize(file\_cv, config$id\_vars, arrow::read\_feather)),  
 tar\_target(df\_cat\_lookup, qract\_read\_and\_anonymize(file\_cat\_lookup, config$id\_vars, arrow::read\_feather)),  
 tar\_target(df\_feat\_lookup, qract\_read\_and\_anonymize(file\_feat\_lookup, config$id\_vars, readr::read\_csv))  
)

## Established \_targets.R and \_targets\_r/targets/load.R.

#### 2.2.2.3 Report

[link](01_data.html)

tarchetypes::tar\_render(  
 report\_data,  
 "src/Rmd/\_01\_data.Rmd",  
 output\_file = "01\_data.html",  
 output\_dir = "src/Rmd/",  
 output\_yaml = "src/Rmd/\_site.yml"  
)

## Established \_targets.R and \_targets\_r/targets/report\_data.R.

### 2.2.3 Time Series Cross Validation

list(  
 tar\_target(p\_tscv, qract\_plot\_tscv(df\_mm, df\_cv, config$max\_year)),  
 tar\_target(df\_cv\_preds\_and\_coefs, qract\_pred\_cv(df\_mm\_bin, df\_cv, df\_form, config$id\_vars))  
)

## Established \_targets.R and \_targets\_r/targets/cv.R.

#### 2.2.3.1 Report

[link](02_cv.html)

tarchetypes::tar\_render(  
 report\_cv,  
 "src/Rmd/\_02\_cv.Rmd",  
 output\_file = "02\_cv.html",  
 output\_dir = "src/Rmd/",  
 output\_yaml = "src/Rmd/\_site.yml"  
)

## Established \_targets.R and \_targets\_r/targets/report\_cv.R.

### 2.2.4 Performance

list(  
 tar\_target(df\_perf, qract\_perf(df\_cv\_preds\_and\_coefs, config$max\_year)),  
 tar\_target(df\_calib, qract\_lin\_calib(df\_cv\_preds\_and\_coefs, min\_sample\_size = 200)),  
 tar\_target(df\_bin, qract\_bin\_preds(df\_cv\_preds\_and\_coefs, n\_bins = 4, confidence\_level = .75)),  
 tar\_target(  
 p\_calib,  
 qract\_plot\_calibration\_pub(  
 category\_id\_str = df\_cv\_preds\_and\_coefs$category\_id %>% unique(),  
 df\_bin,  
 df\_calib,  
 df\_cat\_lookup,  
 uniform\_color = "black",  
 color\_calib = "grey"  
 )  
 )  
)

## Established \_targets.R and \_targets\_r/targets/perf.R.

#### 2.2.4.1 Report

[link](03_perf.html)

tarchetypes::tar\_render(  
 report\_perf,  
 "src/Rmd/\_03\_perf.Rmd",  
 output\_file = "03\_perf.html",  
 output\_dir = "src/Rmd/",  
 output\_yaml = "src/Rmd/\_site.yml"  
)

## Established \_targets.R and \_targets\_r/targets/report\_perf.R.

### 2.2.5 Forest Plots

list(  
 tar\_target(  
 df\_forest,  
 qract\_forest\_plots(  
 df\_cv\_preds\_and\_coefs,  
 df\_feat\_lookup,  
 df\_mm,  
 df\_cat\_lookup,  
 category\_ids = c(  
 "cnsn",  
 "dtin",  
 "sfty",  
 "ptpe"  
 )  
 )  
 ),  
 tar\_target(  
 forest\_files,  
 qract\_save\_forest\_plots(  
 df\_forest,  
 path = "./png"  
 ),  
 format = "file"  
 )  
)

## Established \_targets.R and \_targets\_r/targets/forest.R.

#### 2.2.5.1 Report

[link](%22./04_forest.html%22)

tarchetypes::tar\_render(  
 report\_forest,  
 "src/Rmd/\_04\_forest.Rmd",  
 output\_file = "04\_forest.html",  
 output\_dir = "src/Rmd/",  
 output\_yaml = "src/Rmd/\_site.yml"  
)

## Established \_targets.R and \_targets\_r/targets/report\_forest.R.

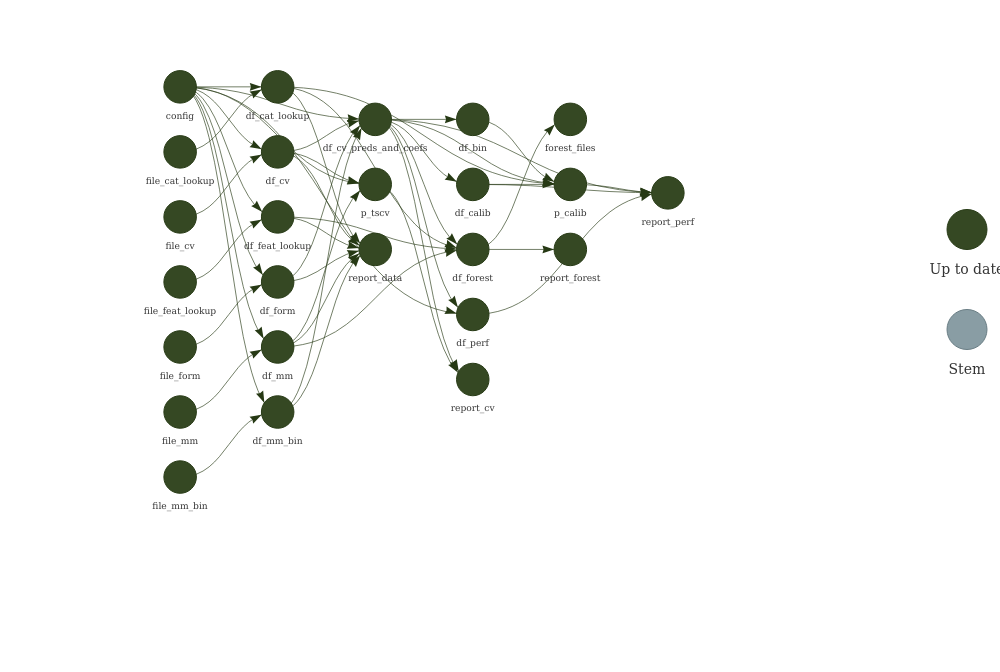
## 2.3 Run Workflow

tar\_make()  
## ✓ skip target file\_mm  
## ✓ skip target file\_form  
## ✓ skip target config  
## ✓ skip target file\_cat\_lookup  
## ✓ skip target file\_cv  
## ✓ skip target file\_feat\_lookup  
## ✓ skip target file\_mm\_bin  
## ✓ skip target df\_form  
## ✓ skip target df\_mm  
## ✓ skip target df\_cat\_lookup  
## ✓ skip target df\_cv  
## ✓ skip target df\_feat\_lookup  
## ✓ skip target df\_mm\_bin  
## ✓ skip target p\_tscv  
## ✓ skip target df\_cv\_preds\_and\_coefs  
## ✓ skip target report\_data  
## ✓ skip target df\_perf  
## ✓ skip target report\_cv  
## ✓ skip target df\_bin  
## ✓ skip target df\_calib  
## ✓ skip target df\_forest  
## ✓ skip target p\_calib  
## ✓ skip target report\_forest  
## ✓ skip target forest\_files  
## ✓ skip target report\_perf  
## ✓ skip pipeline

## 2.4 Visualise Workflow

### 2.4.1 Without Functions

tar\_visnetwork(targets\_only = TRUE, reporter = "silent")



### 2.4.2 With Functions

tar\_visnetwork(reporter = "silent")

