ALLERGY

# EPIDEMIOLOGICAL MODELLING AT THE NEXT LEVEL

15.12.2021

I have done epidemiological modeling for a long time  in SAS and R.  cleaning the data and building an appropriate statistical model. In practice there were basically two choices either logistic or linear regression, maybe in subgroups and defined intervals, with more or less variables depending on experience or preference ( occasionally also  a Loess smoother for a clearer picture). Maybe I learned about R^2 and AIC while writing this paper. Yes, I always looked at residuals but my capability in judging model capability was limited both from theoretical and practical aspects.

When working now on an image analysis project to Python, I found it stunning what the library PyCaret can do.

> One line of code and we have trained and evaluated over 20 models using cross-validation. The scoring grid printed above highlights the highest performing metric for comparison purposes only. The grid by default is sorted using `R2` (highest to lowest) which can be changed by passing `sort` parameter. For example, `compare_models(sort = 'RMSLE')` will sort the grid by RMSLE (lower to higher since lower is better).

I am now thinking to repeat some earlier analysis — just to see if I really did the best choice when writing my latitude paper back in 2005. It was one of the largest studies at that time.

```
conda create --name regression python=3.7.11
conda activate regression
pip install pycaret pyreadr
conda deactivate
```

TBC