

Motoric 6.12.

Fit Model

- do we need to encode categorical variables?

Errors

- measure errors
- which error functions are there? (mean-squared-error)
- which error fits best?
- calculate errors for Input, Prediction
→ are there more steps where we can calculate errors?

Ensembling models

- What is it?
- How does it work?
- Can it be helpful for us?
- How hard is it to implement?

Needed next steps:

- cleaning
- scaling
- imputation
- error
- balancing
- Models
- Hyperparameters

Notes internal presentation

- use "binary classes" instead of "cat 1-5"
 - it doesn't matter if a child is in class 1 or 2, it only matters if it is lacking or not

Problems

- the tipcode in cbs gender age has a one-to-many relationship
 - split Dataframe in 3 Dataframes?
- unique values ID or (gender + age)

- CBS income has only region → region to zipcode
- Regions also have a one to many relationship
- Unique values only ID or Kennmerken Van Huishoudens

CBS Core Numbers

- Postcode is Coding

- Regio = Wijk (Viertel / Stadtteil)
 \rightarrow Wk 168000

Gemeinde - wijkcode / Viertel code
 code

\rightarrow in total 2 Letters + 6 Numbers

- Regio = Buurt (Gegend)

\rightarrow BU 16800003

Gemeinde / Buurtcode / Gegendcode
 Postcode

wijkcode/
 Viertelcode

\rightarrow in total 2 Letters + 8 Numbers

- Den Haag = S-Gravenhage

{ \rightarrow only rows where Regio = Zipcode copy

\rightarrow make smaller DF from CBS Data

Both for this and CBS income useful

CBS Migration background

- zipcode has one to many relationship
- unique values ID + (gender + age)

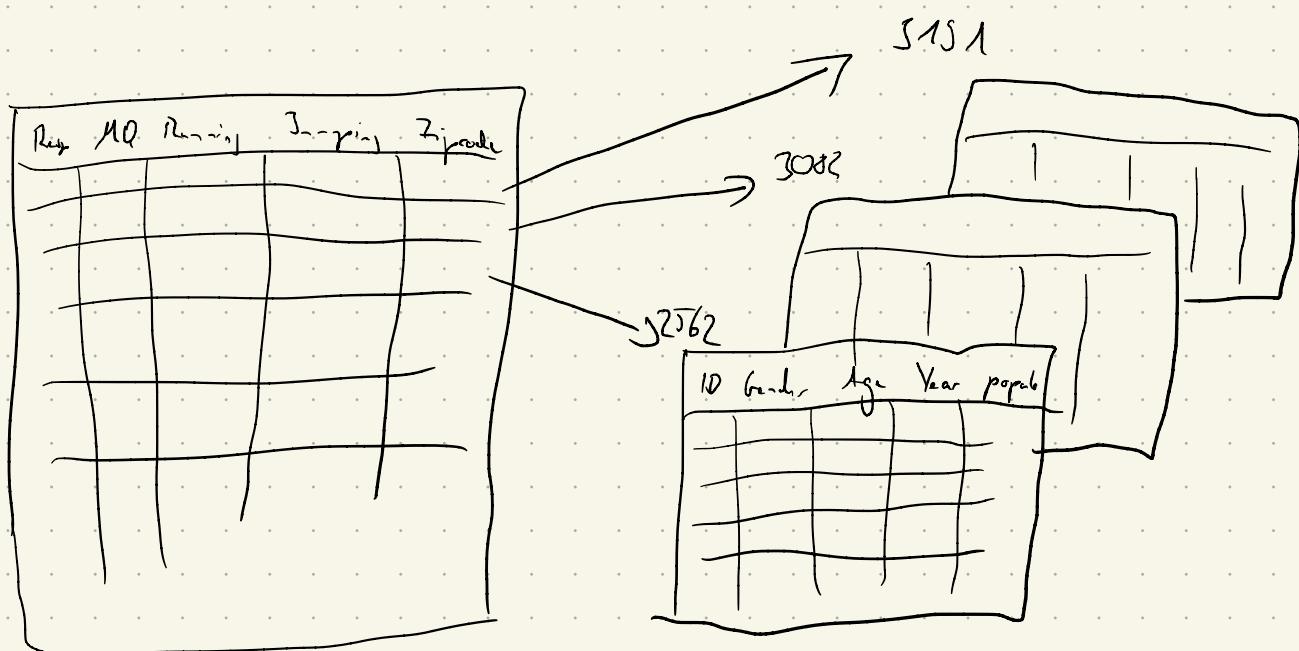
CBS

gender + age

→ zipcode was key word

→ needs Dataframe pro zipcode

→ benötigt alle unique zipcodes aus original df



→ merge columns

$$\rightarrow \text{Total} - 0 - 5 =$$