# Generating models from UTH data and test on NC data.

## Step1: Name2vec

* Generate 2 models for first\_name and laste\_name.
* input data --> NC + ONC data
* code --> H:\mahin\thesis\name2vec
* run this on ViDaL LINUX and get the 2 models.

## Step2: Feature extraction - UTH data

* code --> code\1\_feature\_extraction.R
* result --> data\df\_train\_all.rds & data\df\_test\_all.rds
  + all features:
    - dob
    - fname
    - lname
    - mname
    - name swap
    - gender
    - mariage
    - SSV
    - address
    - phone

## Step3: add Name2vec using models from step1 - UTH data

* code --> name2vc\add\_feature\_name2vec.py
* input (from step2) --> name2vc\train.csv & name2vc\test.csv
* result --> name2vec\train\_n2v.csv & name2vec\test\_n2v.csv

## Step4: name2vec post processing

* code --> code\2\_n2v\_postprocess.R
* input --> name2vec\train\_n2v.csv & name2vec\test\_n2v.csv
* result --> data\df\_train\_all\_n2v.rds & data\df\_test\_all\_n2v.rds

## Step5: models

* code files --> code\3\_\*\_model\_\*.R
* result -> result\

## Step7: Generate pairs – NC data

* code 🡪 H:\mahin\autoRL-NCV-master\
* result 🡪 H:\mahin\autoRL-NCV-master\autoRL-NCV-master\data\generated\NC\_pair.csv

## Step6: Feature extraction - NC data

* code 🡪 code\ 4\_NC\_feature extraction.R
* result 🡪 data\NC\_test\_all.rds & name2vec\NC\_test.csv

## Step7: add Name2vec using models from step1 - NC data

* code --> name2vc\add\_feature\_name2vec.py
* input (from step6) --> name2vec\NC\_test.csv
* result --> name2vec\NC\_test\_n2v.csv

## Step8: name2vec post processing – NC data

* code --> code\2\_n2v\_postprocess.R
* input --> name2vec\ NC\_test\_n2v.csv
* result --> data\NC\_test\_all\_n2v.rds

## Step9: Test – NC data

* code 🡪 code\5\_test\_NC.R