

IRONHACK DATA ANALYTICS BOOTCAMP
FINAL PROJECT

PREDICT: ARE YOU SATISFIED WITH YOUR LIFE?

A machine learning approach from Elvira Hänni

QUESTIONS

SATSIFACTION WITH LIFE

- What features impact satisfaction with life?
- How ist it possible to predict satisfaction with life?



DATASET

EUROPEAN UNION STATISTICS ON INCOME AND LIVING CONDITIONS

Syntetic Data for Germany - 22'000 persons



Background:

- Year of birth
- Sex
- Marital status
- ...



Job:

- Economic Status
- Type of job
- Managerial position
- ...



Basic needs:

- Two pairs of shoes
- Internet connection
- Have clothes
- ...



Education:

- Education activity
- Highest level



Income:

- Salary
- Social benefits
- ...



Health:

- General health
- Need medical examination
- ...

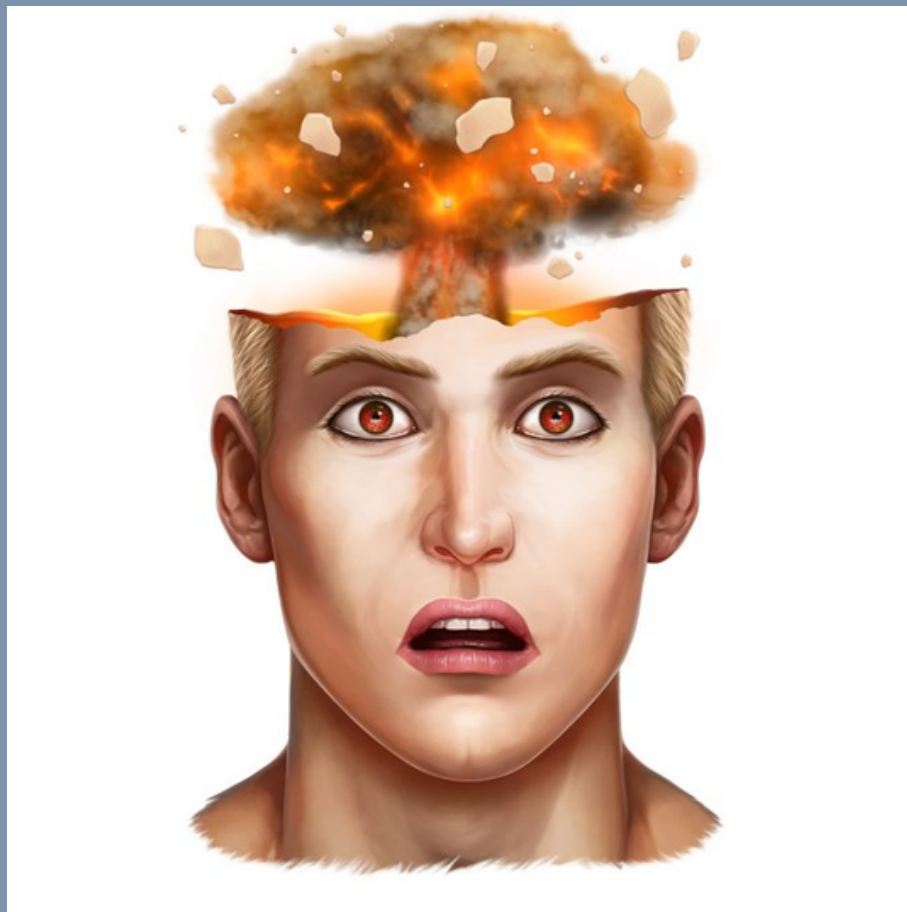


Satisfaction with...:

- life overall
- political system
- ...

CHALLENGES

A lot of datas



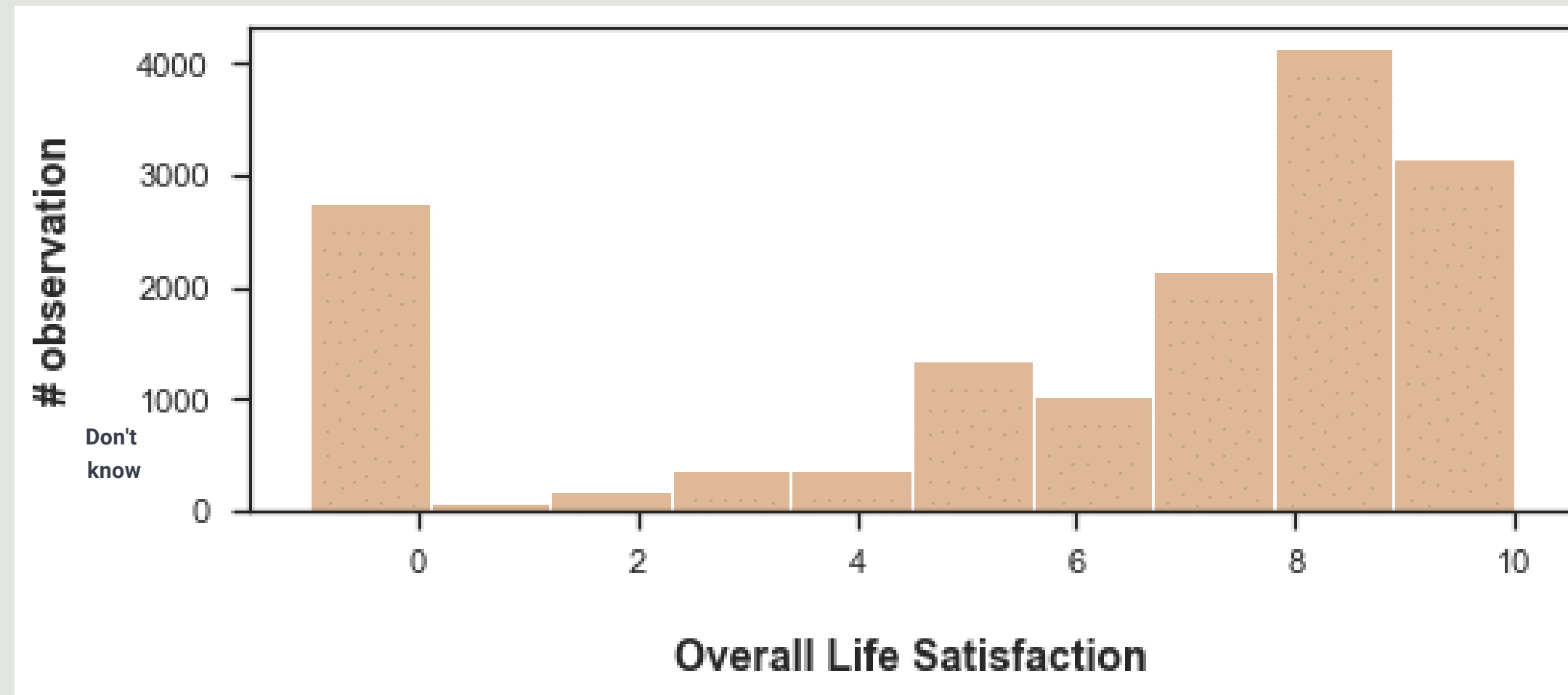
- a lot of missing values
- a lot of "don't know"
-

Find a model to build on

1	<code>r2_score(y_test, y_pred)</code>
-0.0010766855291548527	

	Satisfied	Not satisfied
True	2621	0
False	1272	0

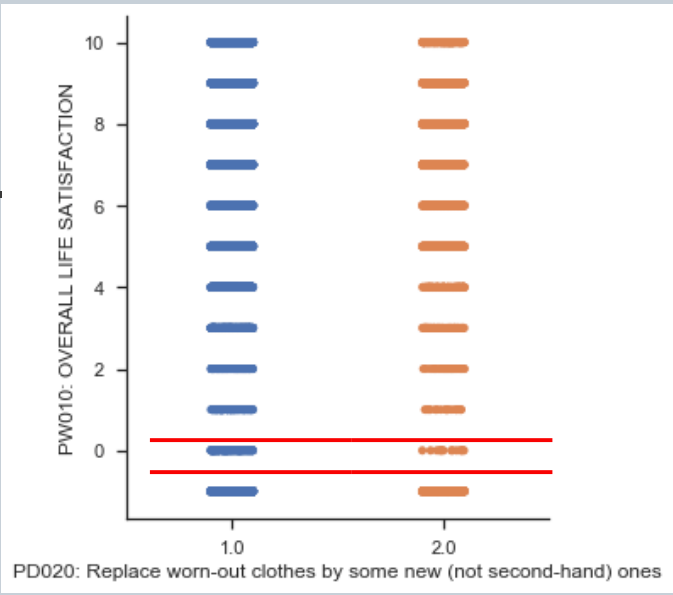
WHAT IMPACTS SATISFACTION WITH LIFE?



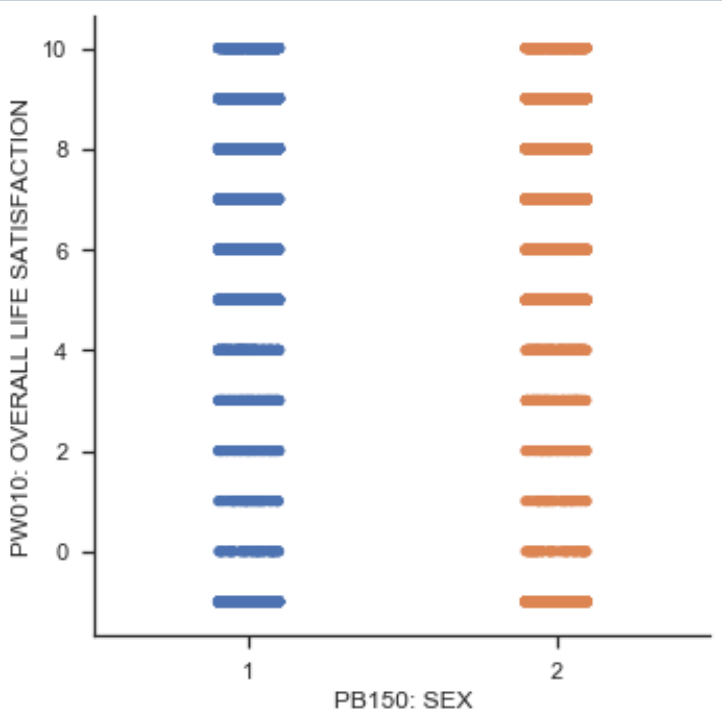
WHAT IMPACTS SATISFACTION WITH LIFE?

RECURSIVE FEATURES SELECTION

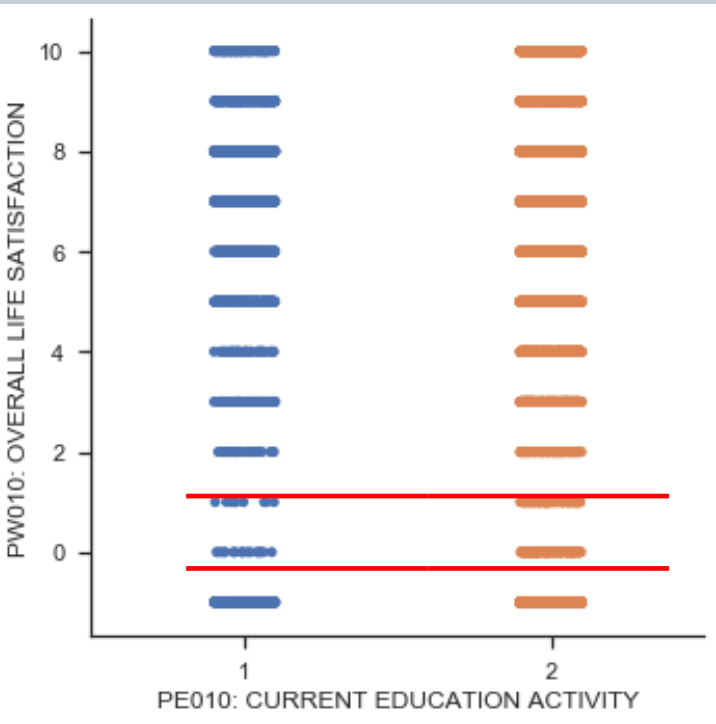
Did you replace worn-out clothes?



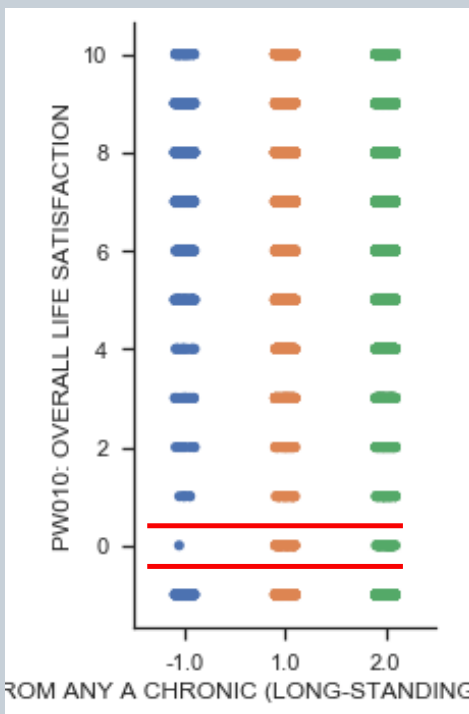
Sex



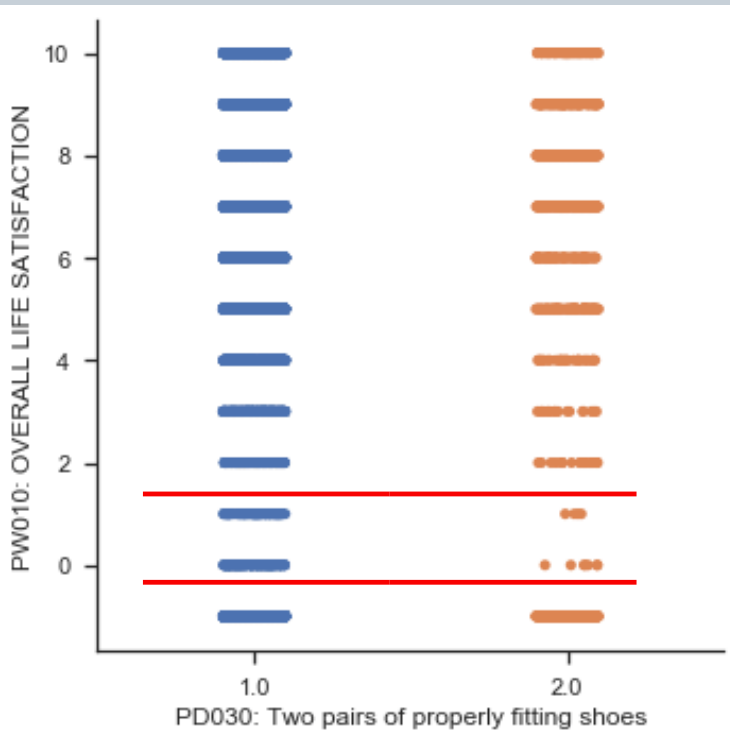
Current Education Activity



Suffer from any chronic illness



Do you have two pairs of properly fitting shoes?



MODEL TO PREDICT SATISFACTION WITH LIFE

REGRESSION

1	r2_score(y_test, y_pred)
<u>-0.0010766855291548527</u>	

CLASSIFI-
CATION

	Satisfied	Not satisfied
True	2612	0
False	622	0

MULTI-
CLASSIFI-
CATION

START

TOOLS

APPLYED IN SUPERFUNCTIONS



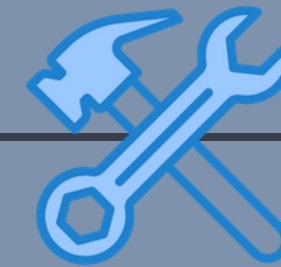
PREPARE DATA

- Drop nan /dont know, change nan
- Binning y and X
- Create Dummy
- Change Features and target based on RFE and others



ML PREPERATION

- Scaling (...)
- PCA / ICA



METHODS

- 4 Regressionmodels
- 4 Classification models
- 2 Multiclassmodels



HYPERTUNING

- Taylor Polynomial
- Deciscion Tree Classifier

Pipeline

MODEL TO PREDICT SATISFACTION WITH LIFE

REGRESSION

$R^2 = 0.00051$

all the other neg R^2

CLASSIFI-
CATION

Accuracy: 57%

Precision: 31%

Recall: 69%

MULTI-
CLASSIFI-
CATION

CONCLUSION AND NEXT STEPS

MAYBE SATISFACTION WITH LIFE IS JUST NOT PREDICTABLE?



TECHNICAL STEPS

- Try with other dataset (real dataset, other country / year)
- Underfitting, Balancing



CONTENT-RELATED STEPS

- Trust in politics
- Trust in legal systems