Predicting life expectancy

Project supervised machine learning, 2nd October 2019

Filipa Lopes

Motivation

- Insurance companies
- Retirement provision

10 Predictors

Target: life expectancy

(~2000 entries)

Your:

- yearly income*
- CO_2 emissions (tones/person)
- total spending in healthcare*
- Allowance/ government spending in your healthcare*
- food consumption (calories/day)
- sugar consumption (grams/day)
- schooling (years spent on school + university)
- alcohol consumption
- body mass index (BMI)
- fertility (number of children)

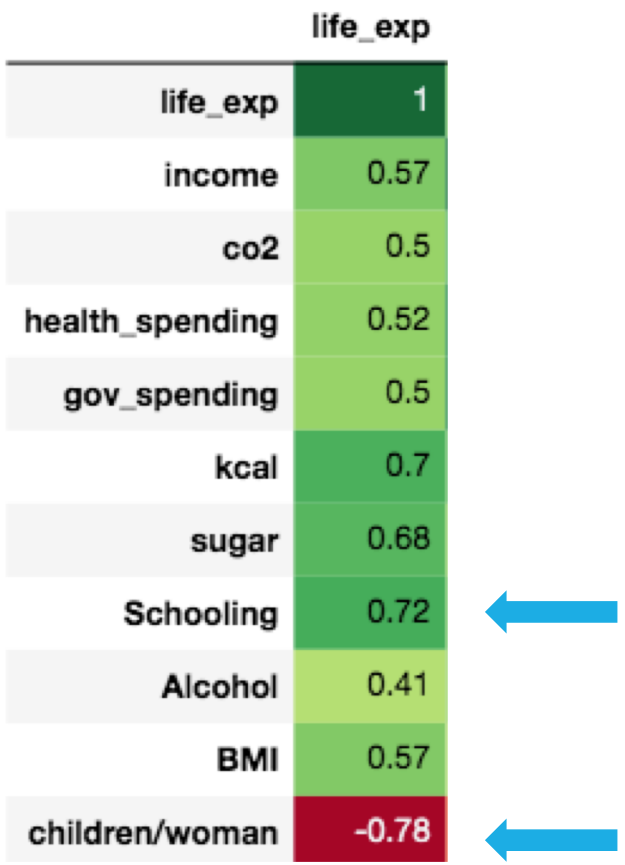
*($US, inflation-adjusted)

Data

- 187 countries, from 2000 till 2010
- ~2000 entries
- Sources: World Health Organization(Kaggle) and gapminder.org

What are the best predictors for life expectancy?

Correlation heatmap

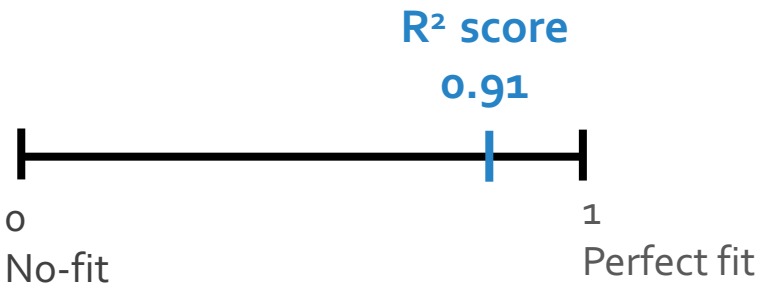


Models

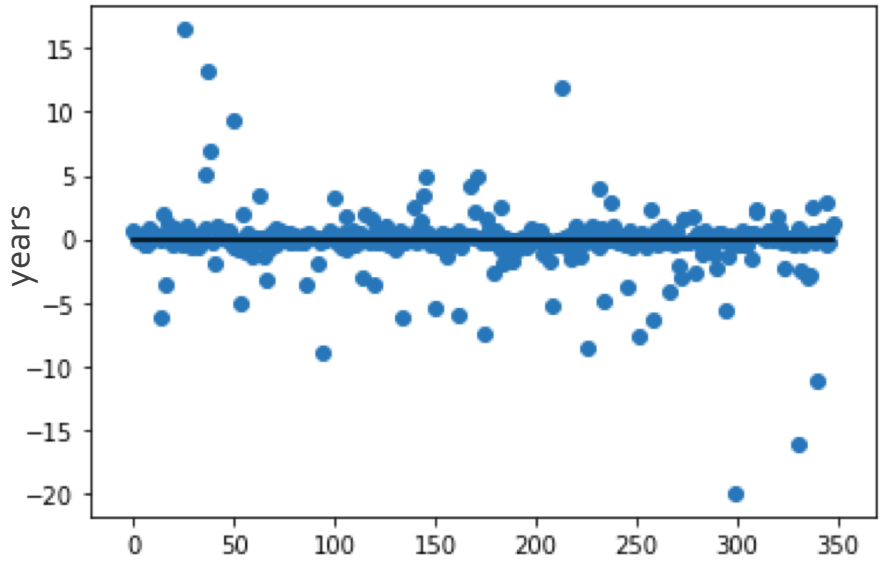
Test accuracy

- Linear Regression model: 76.64%
- KNRegression (neighbors=3): 72.04%%
- **Decision Tree model: 91.33%**

Evaluation of model



How much the model does not match?



(observed values – predicted values)

Real case scenario

Conclusion (http://filipamiralopes.pythonanywhere.com)