Physical activity and health satisfaction of Estonian school workers

MTAT. 03.227 Machine Learning
Mooses K, Pajusalu M, Panov L, Rebecq V
PROJECT OWNER

UNIVERSITY of TARTU
Institute of Sport Sciences and Physiotherapy

Project "Supporting the physical activity of school children"
DATA OVERVIEW

PERIOD
September 2019
December 2019

SUBJECTS
44 schools from 14 counties
888 subjects

JOB POSITIONS
Administrative 17.8%
Teachers 82.2%

VARIABLES
Sociodemographic (gender, age, profession, height, weight, BMI, school and county)
Activity (time of doing sports, number of active days, active transport taken, different intensities)
Satisfaction (health and job satisfaction)
DATA OVERVIEW

PASSPORT OF AVERAGE SUBJECT

Lambda women
- 46 years
- 71 kg
- 167 cm
- BMI: 25.4

Lambda men
- 42 years
- 87 kg
- 181 cm
- BMI: 25.6

AVERAGE MINUTES

Sedentary activity
- 592 min

Light activity
- 213 min

MVPA (moderate to vigorous)
- 50 mins

Women 87.8%
Men 12.2%
STUDY QUESTION
PREDICT SATISFACTION WITH HEALTH
STUDY QUESTION
PREDICT SATISFACTION WITH HEALTH
PREDICT SATISFACTION WITH HEALTH
SATISFACTION WITH HEALTH BY GENDER AND AGE
**DECEMBER**

Testing different methods and combination of features to predict satisfaction with health

- KNN classifier
- Decision tree classifier
- Logistic regression
- Oversampling

**Ensemble classification:**
- Random forest
- Majority Vote

Combining and interpreting the results

**NOVEMBER**

- Data
- Descriptive statistics
- First attempts to ML
- Individual work and discussions

80% of subjects are satisfied with health – few unsatisfied
### BEST RESULTS

#### SATISFACTION WITH HEALTH – TOP 3 MODELS:

<table>
<thead>
<tr>
<th>PRECISION</th>
<th>RECALL</th>
<th>VARIABLES</th>
<th>METHOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>62%</td>
<td>Age, BMI, %MVPA</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>86%</td>
<td>83%</td>
<td>BMI</td>
<td>Decision Tree</td>
</tr>
<tr>
<td>84%</td>
<td>88%</td>
<td>Age, gender, BMI, %MVPA, %Light PA, %Sedentary</td>
<td>Random Forest</td>
</tr>
</tbody>
</table>

We can also predict:
- job position (89%)
- where does the subject come from (school 92%, county 97%).
When predicting the satisfaction with health, physical activity data was not very specific, especially for detecting 'not satisfied'
LESSONS LEARNED

- Although rather small and unbalanced dataset, it still enabled to find satisfactory models.
Thank you!

Questions?

Project in GitHub