paper-summary

Data

Sourced from Colombia, Peru, Mexico

23% of data collected directly from anonymous users with web survey (485 total)

77% of data generated artificially with the Weka tool and SMOTE filter

# of attributes: 17

# of records: 2111

Goal: identify obesity level of an individual and monitor obesity levels using a recommender system trained using this data-set

Attributes

Related to Eating Habits:

- FAVC (Frequent consumption of high caloric food)
- FCVC (Frequency of consumption of vegetables)
- NCP (Number of main meals)
- CAEC (Consumption of food between meals)
- CH2O (Consumption of water daily)
- CALC (Consumption of alcohol)

Related to Physical Condition:

- SCC (Calories consumption monitoring)
- FAF (Physical activity frequency)
- TUE (Time using technology devices)
- MTRANS (Transportation used)

Other

- Gender
- Age
- Height
- Weight
- Family History of Obesity

Goal Attribute

NObesity

Calculated using mass body index (Weight/Height^2)

Classified into following categories:

- Insufficient Weight

- Normal Weight
- Overweight Level I
- Overweight Level II
- Obesity Type I
- Obesity Type II
- Obesity Type III

Issues

Data initially unbalanced as categories not represented equally. (see Fig. 1)

Problematic for data mining methods.

More accurate for categories with more data, and vice versa.

Solution: Generation of synthetic data with Weka tool and SMOTE filter

Personal Remarks

- diff diets, lifestyles, what is "normal"?
- this summary can serve as a introduction to the data set
- i dont really understand the processing w/ smote but i believe it is irrelevant for our purposes