


MSc thesis topics
Explanatory Data Analysis

Matthijs van Leeuwen
Leiden Institute of Advanced Computer Science

27 June 2019



Universiteit
Leiden

We do research on



Exploratory data mining

- Pattern mining
- Interactive data mining
- Interpretable mining & machine learning
- Information theoretic data mining

Interdisciplinary data science

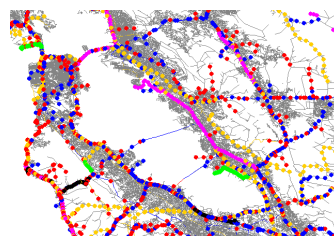
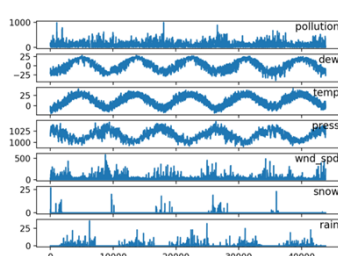
- Social sciences
- Life sciences
- Industry
- Law



Example topics related to Information Theoretic Data Mining

Clustering time series data by
compression

Compression-based pattern mining
from spatiotemporal graphs



Dementia back in the heart of society



- 1) Visualization of human movement trajectories
- 2) Automatic extraction of interaction moments
from human movement trajectories



Thesis

MDL-based approach

Research project

Data collection for validation

A novel algorithm for p-privacy in medical data



Privacy very important for data publication

Develop a generalisation algorithm that optimises **p-privacy** while minimising **loss**



Instance-based explanation of predictions by machine learning models

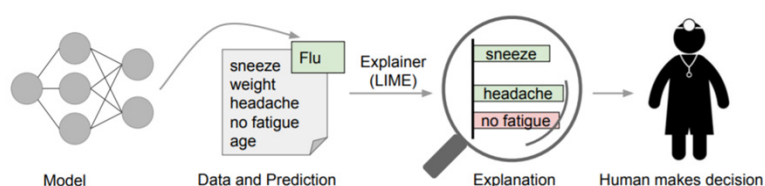


Explainable machine learning

Use instances from data as **exemplars**

Model-agnostic

Formalise problem and develop new algorithm



Optimising probabilistic rule lists using evolutionary algorithms



Probabilistic rule lists are considered interpretable machine learning models

```

IF {backbone = no} THEN Pr(invertebr.) = 0.55
                                Pr(bug) = 0.45
ELSE IF {breathes = no} THEN Pr(fish) = 0.93
                                Pr(reptile) = 0.07
ELSE IF {feathers = yes} THEN Pr(bird) = 1.00
ELSE IF {milk = no} THEN Pr(reptile) = 0.50
                                Pr(amphibian) = 0.50
ELSE THEN Pr(mammal) = 1.00
  
```

Idea: use **evolutionary algorithms** to find **better rule lists**

Possible interests

Exploratory data mining
Explainable machine learning
Interactive data analytics
Interpretability
Privacy

Your skills

Algorithmics
Data mining
Data structures
Experimentation
Programming

More information (to be completed soon):

eda.liacs.nl/for-students

