Occupancy Detection

MT7038

Anton Stråhle & Max Sjödin

Fall 2020

Data

The occupancy status of a room was observed for a few days. Snapshots of the features below were taken every minute.

- Features - Numerical
  - Temperature
  - CO2
  - Humidity
  - HumidityRatio
  - Light
- Labels - Binary
  - Occupancy
    - Occupied
    - Unoccupied

Data

The occupancy status of a room was observed for a few days. Snapshots of the features below were taken every minute.

- Features
  - Temperature
  - CO2
  - Humidity
  - HumidityRatio
  - ~~Light~~
- Response
  - Occupancy
    - Occupied
    - Unoccupied

Light is excluded as the best classifier would otherwise become *Are the lights on?*

Brief Exploration

- Unbalanced data set
  - Many more unoccupied data points than occupied

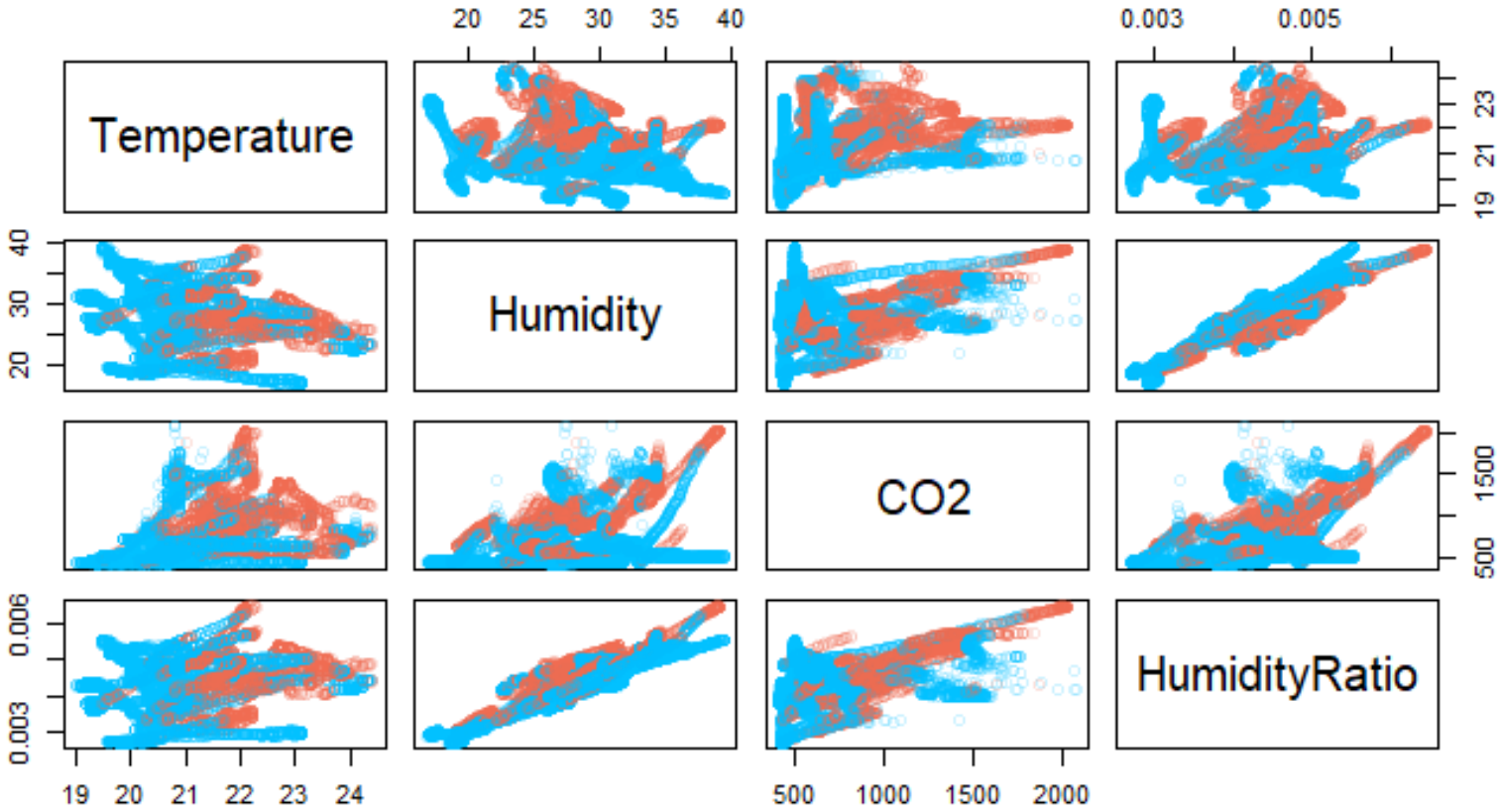


Figure: Pairplots of Features

- Non-linearity?

Brief Exploration

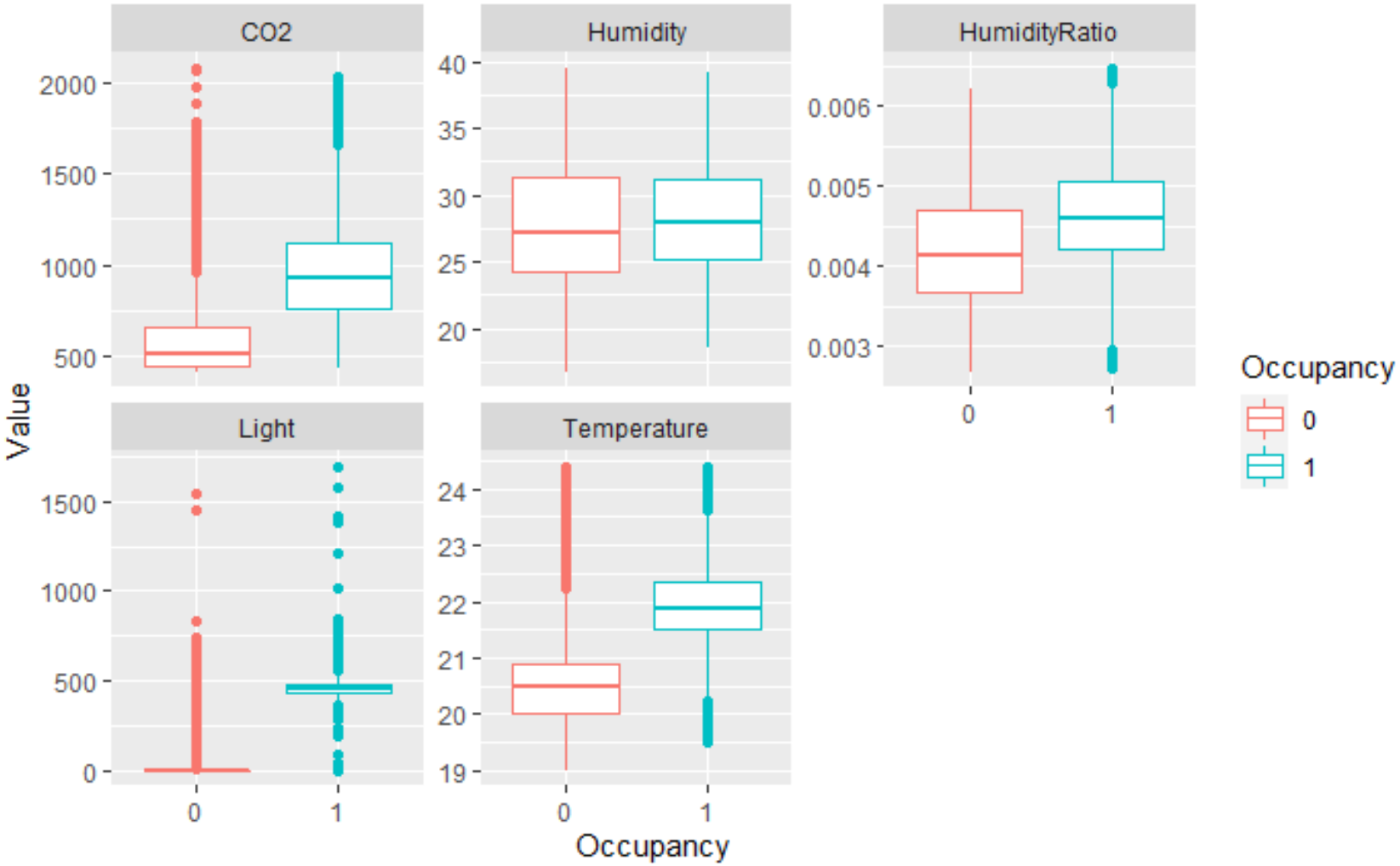


Figure: Boxplots of Features

Methodology

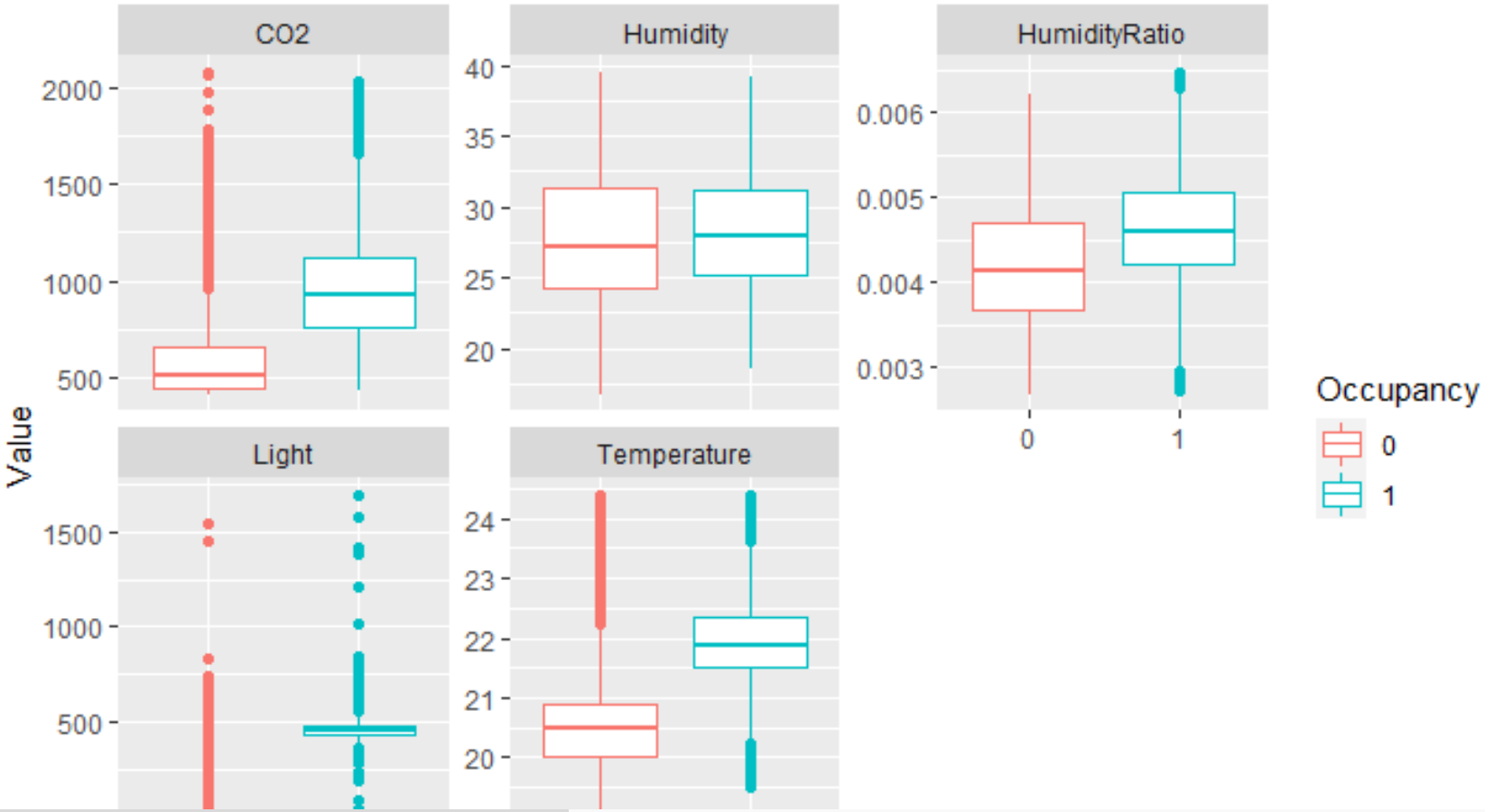
- SVM
- KNN
- Decision Trees

Methodology

SVM

Why? ▷ Good for non-linear classification problems.

How? ▷ Using the package **e1071** and the function **svm**
▷ Linear, polynomial and radial kernels
▷ Coarse-to-fine parameter search



Methodology

KNN

Why?
- ▷ Good for non-linear classification problems
- ▷ Good with large training data sets
- ▷ Good if data is not noisy

How?
- ▷ Regular KNN using the package **class** and the function **knn**
- ▷ Weighted KNN using the package **kknn** and the function **kknn**

Methodology

Decision Trees

Why? ▷ WHY???
▷ WHY MORE?

How? ▷ HOW?
▷ HOW MORE?