University of Waterloo
Electrical & Computer Engineering Department
ECE 659 – Intelligent Sensors and Sensor Networks

Occupancy Detection Through Light, Humidity, CO2 and Temperature Measurements

Presentation prepared by: Yerbol Aussat

Why Occupancy Detection in Buildings?

- ENERGY SAVINGS
  - ~40 % heating energy savings with appropriate management of HVAC systems with occupancy detection (Candanedo et al, 2016)
  - ~30 % lighting energy savings from smart lighting control (Liu et al, 2016)

- SECURITY

- OCCUPANT BEHAVIOR

Typical Approach for Sensing Occupancy

- Passive Infrared Sensors (PIR)
  - Sensor may be triggered by air currents or fail to detect presence when occupant does not move
  - Not reliable for all applications

- Digital Cameras
  - Privacy concerns

Alternative Approach

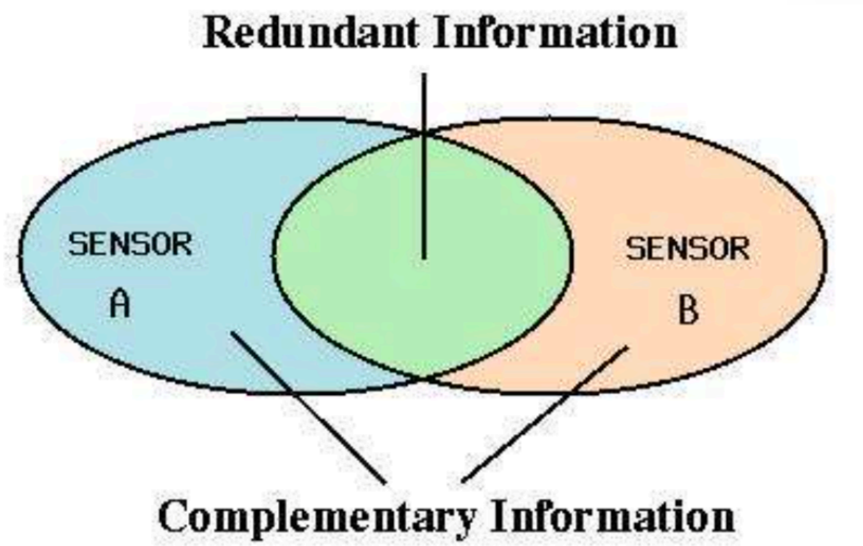
- Fusing data from various low-cost sensors:
  - Temperature
  - Humidity
  - CO_2
  - Light

- Combinations of these sensors can already be found in many buildings

Data Fusion

- Synergistic combination of information made available by various knowledge sources such as sensors, in order to provide a better understanding of a given scene. (Hall, 1992)

- 3 Levels:
  - **Sensor Fusion**
  - **Feature Fusion**
  - Decision Fusion



* Image was taken from lecture slides (ECE659, Uwaterloo, Spring 2017)

Dataset Description

- UCI Data Repository:
  - Occupancy Detection Data set
  - http://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+
- Measured Data:

Type of Data	Description / Units
Time Stamp	Year–Month–Day Hour:Minute:Second
Temperature (T)	In Celsius
Relative Humidity (Phi)	In %
Light	In lux
CO_2	In ppm
Occupancy	0 or 1, (ground occupancy from digital camera)

Data Pre-Processing. Feature Extraction

- **Humidity Ratio (W)**
- Derived from temperature and relative humidity

$$W = 0.622 \frac{p_w}{p - p_w}$$

p - standard atmospheric pressure

p_w – saturation pressure over liquid water (function of temperature and relative humidity)

Data Pre-Processing. Feature Extraction

- Two features were extracted from timestamp:

- **Seconds from Midnight (SM)**
  - 3600*hour + 60*minute + second

- **Week Status (WS)**
  - 0: Weekend
  - 1: Weekday

Training and Testing Datasets

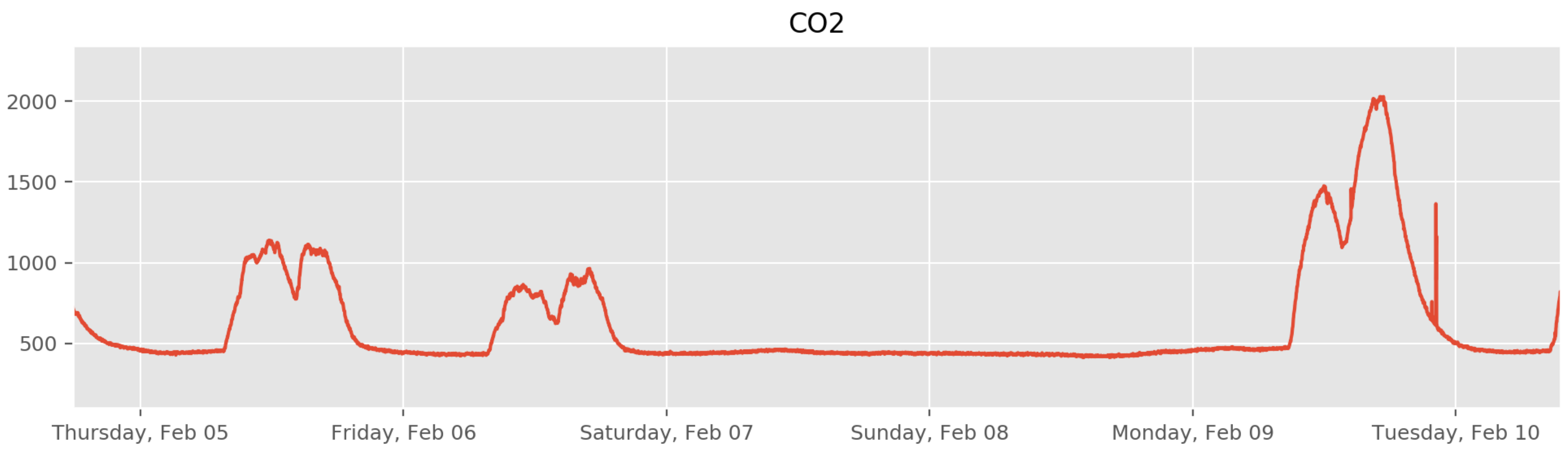
- Total number of samples: 20390

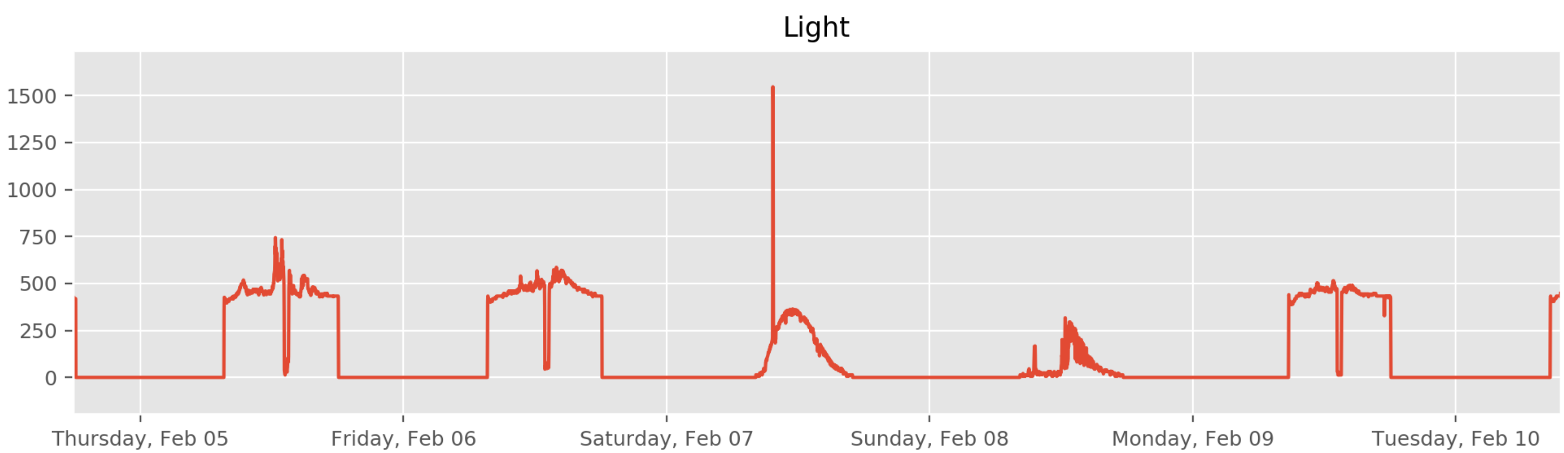
Data Set	Number of samples	Occupancy Data Distribution		Comment: During the Occupied status:
		0 (Not occupied)	1 (Occupied)	
Training	8143	0.79	0.21	Door closed
Testing 1	2665	0.64	0.36	Door closed
Testing 2	9752	0.79	0.21	Door open
Total	20390	0.67	0.23	

Tools Used

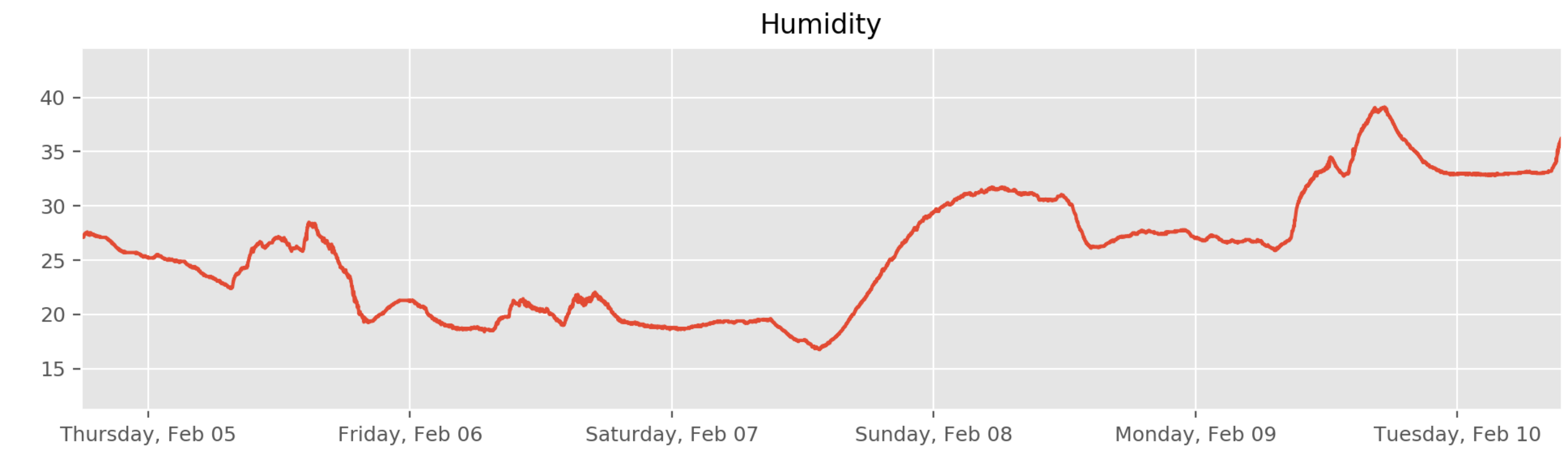
- Pandas: Python Data Analysis Library
- Scikit-Learn: Python Machine Learning Library
- Matplotlib: Python 2D plotting Library

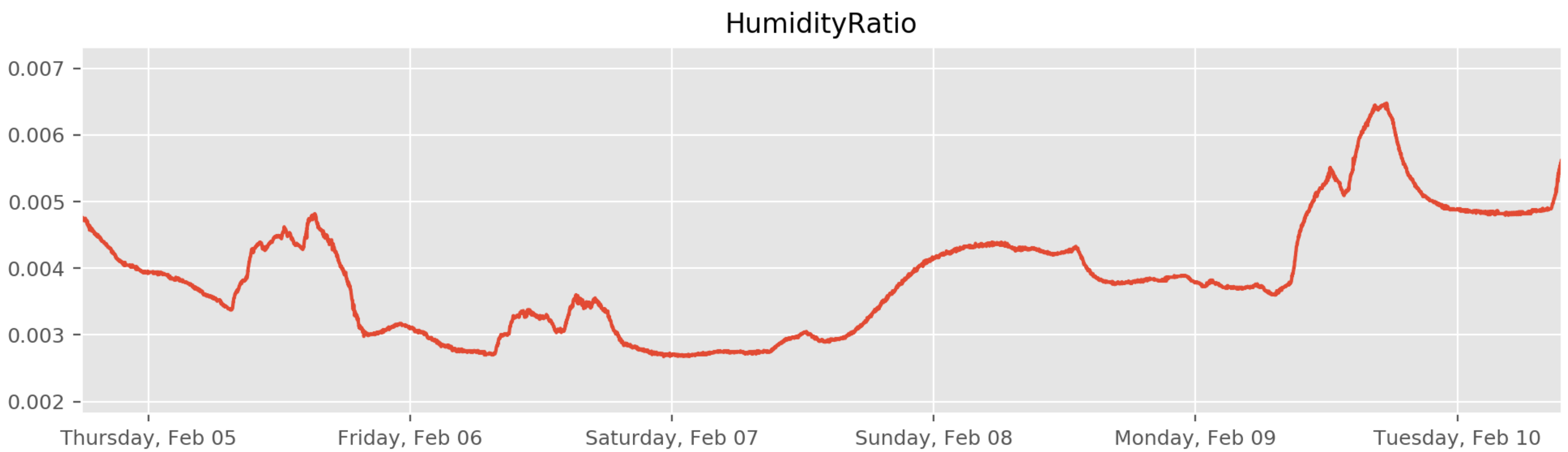
Plots for Individual Signals



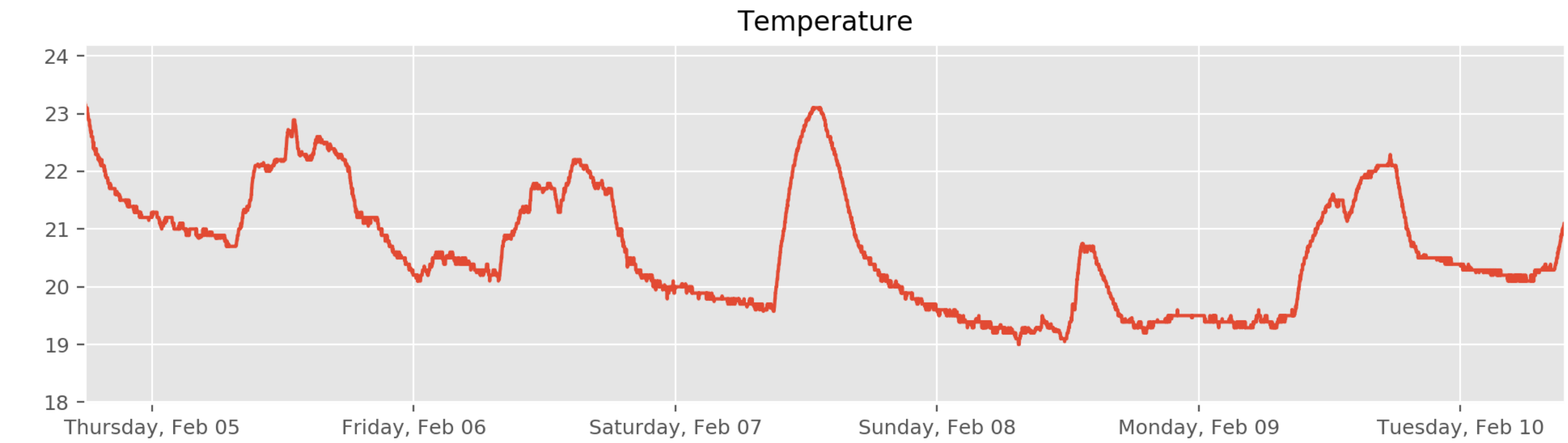


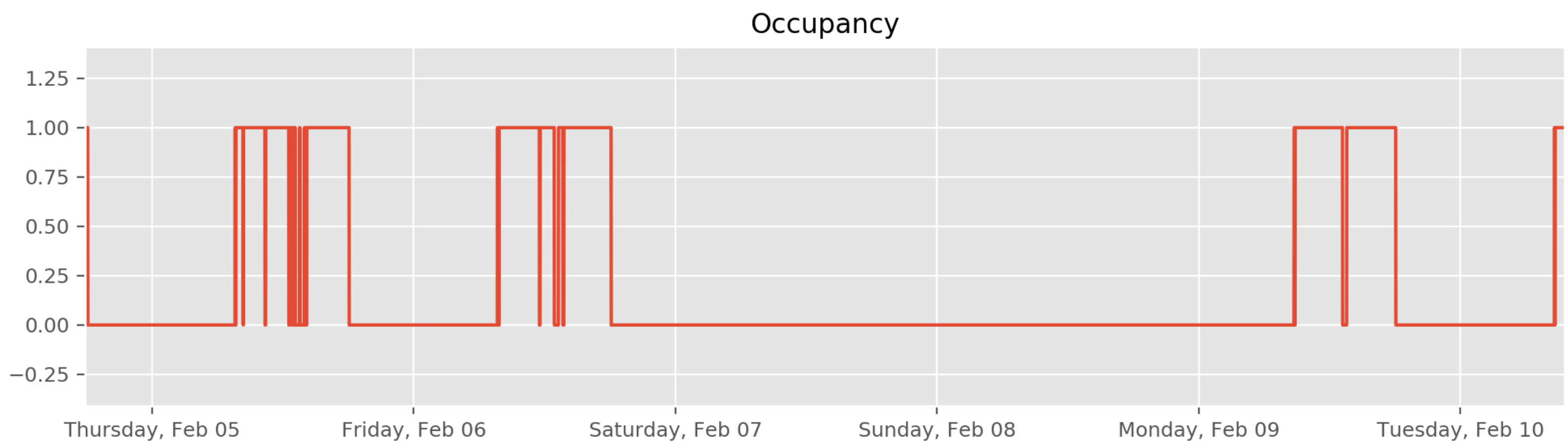
Plots for Individual Signals





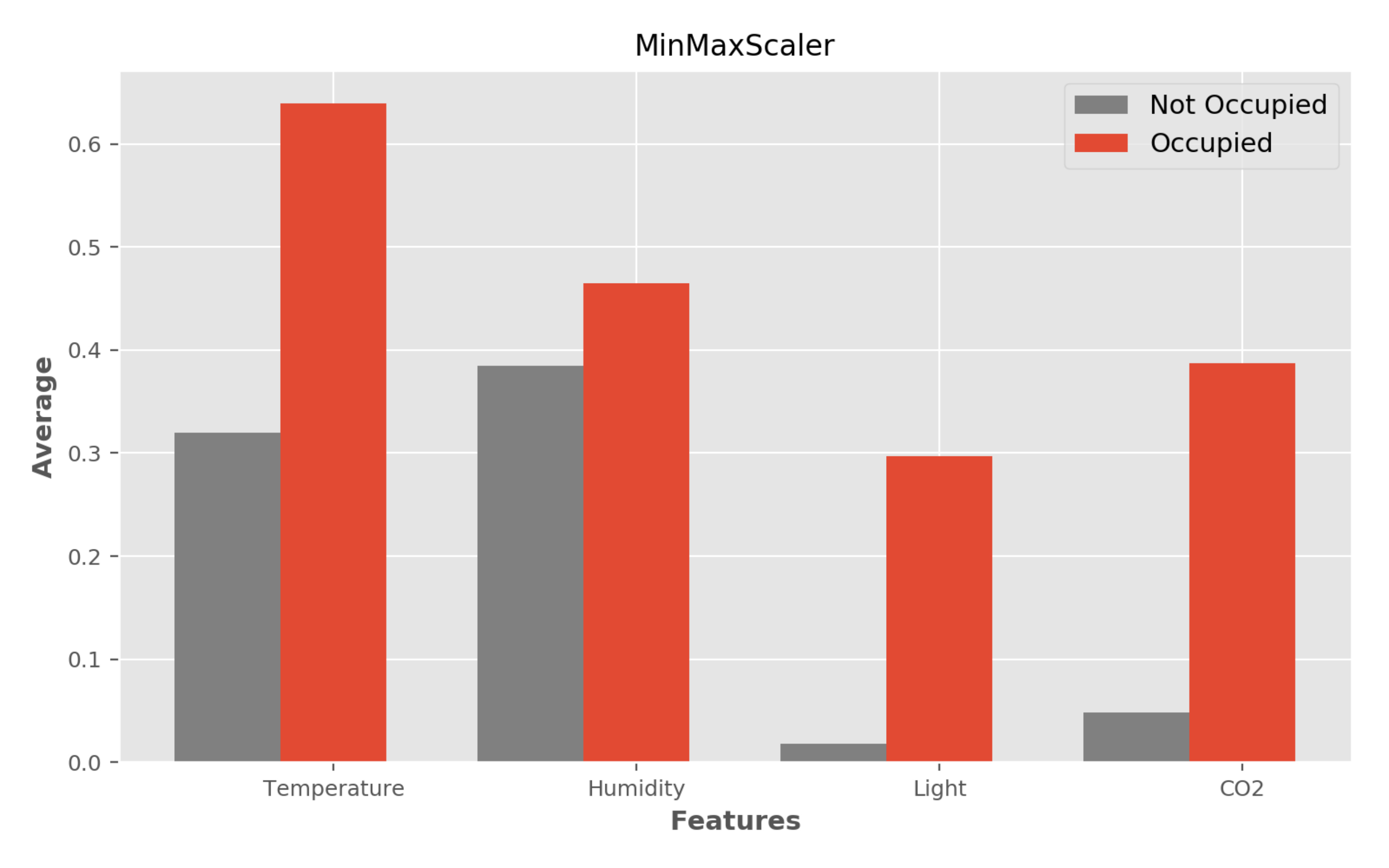
Plots for Individual Signals





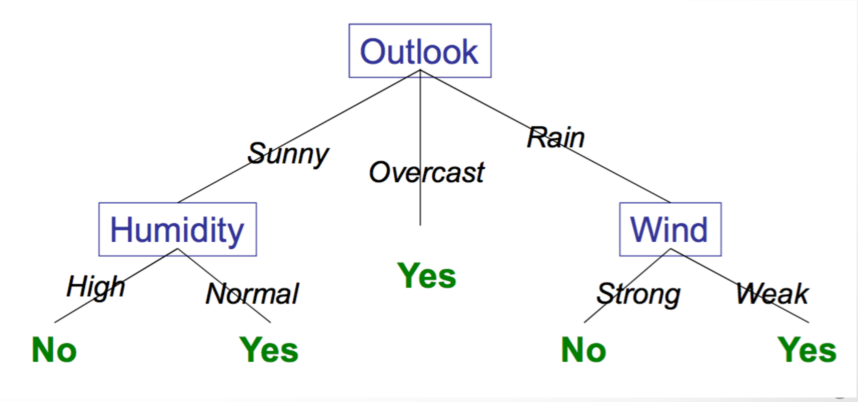
Scaled Individual Signals

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$



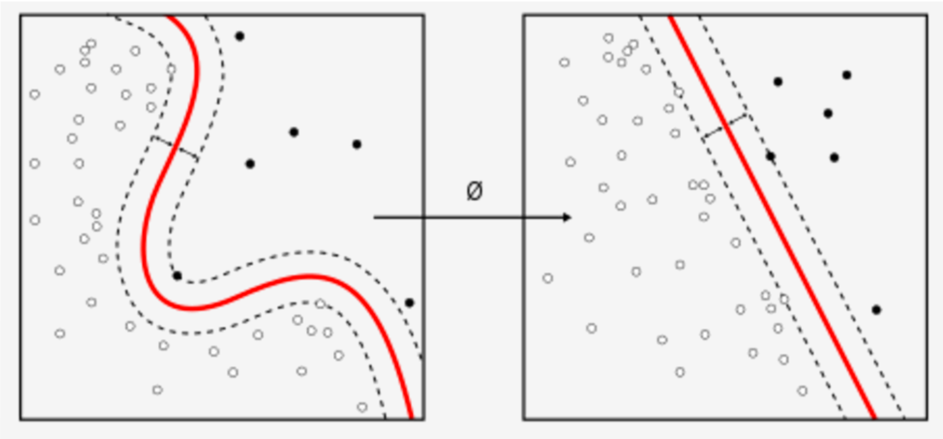
Data Fusion Techniques Used

- Naïve Bayes (NB)
  - Probabilistic Data Fusion Method
  - Based on Bayes' theorem
  - Strong independence assumption between features
- Decision Trees (DT)
  - AI-based Data Fusion Method
  - Classification trees (Discrete set of values)
  - Leaves represent labels
  - Nodes represent features
  - Branches represent values of features



Fusion Techniques Used

- Random Forrest (RF)
  - AI-based Data Fusion Method
  - Constructs a large collection of correlated decision trees
  - The idea is to average noisy and unbiased models to create a model with low variance
  - Correct for decision trees' tendency to overfit
- Support Vector Machines (SVM)
  - Non-probabilistic binary classifier
  - Represents examples as points in space, mapped so that the examples of different categories are divided by a gap
  - New examples are then mapped into the same space and predicted to belong to a category based on which side of the gap they fall



Evaluation Model: Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

Correctly Classified Ratio (Accuracy Rate):

$$\%CCR = \frac{a+d}{a+b+c+d} * 100\%$$

Preliminary Results

Model	Parameters	Training Accuracy (%)	Testing Accuracy 1 (%)	Testing Accuracy 2 (%)
NB	T, Phi, L, CO2, W, WS, SM	97.888	**97.749**	**98.554**
RF	T, Phi, L, CO2, W, WS, SM	100.000	**94.747**	**97.303**
DT	T, Phi, L, CO2, W, WS, SM	99.988	**94.522**	**91.981**
SVM	T, Phi, L, CO2, W, WS, SM	100.000	63.565	78.989
NB	L, WS, SM	98.060	97.073	94.114
RF	L, WS, SM	99.975	**97.073**	**97.908**
DT	L, WS, SM	99.828	96.510	96.975
SVM	L, WS, SM	99.939	71.595	82.927
NB	Phi, CO2, WS, SM	78.767	63.527	78.989
RF	Phi, CO2, WS, SM	99.963	**95.422**	**79.984**
DT	Phi, CO2, WS, SM	99.939	**95.310**	64.858
SVM	Phi, CO2, WS, SM	100.000	63.640	79.020

Preliminary Results

Model	Parameters	Training Accuracy (%)	Testing Accuracy 1 (%)	Testing Accuracy 2 (%)
NB (Best)	L	97.728	**97.711**	**98.913**
NB (Best)	CO2	90.225	87.280	78.640
NB (Best)	T	84.195	84.878	85.131
NB (Best)	Phi	78.964	63.527	78.640
NB	WS, SM	78.767	63.527	78.989
RF	WS, SM	97.347	**92.296**	**92.135**
DT	WS, SM	97.347	**93.208**	**94.180**
SVM	WS, SM	96.721	**91.520**	**94.053**
NB	Phi, CO2	89.623	86.979	79.061
RF	Phi, CO2	99.853	80.300	42.504
DT	Phi, CO2	99.067	79.099	38.167
SVM	Phi, CO2	98.391	74.747	41.981

Preliminary Results

- Using normalized features (minmax scaler)

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Model	Parameters	Training Accuracy (%)	Testing Accuracy 1 (%)	Testing Accuracy 2 (%)
NB	Phi, CO_2, WS, SM	80.965	70.769	79.081
RF	Phi, CO_2, WS, SM	99.988	94.484	75.236
DT	Phi, CO_2, WS, SM	99.939	91.707	66.048
SVM	Phi, CO_2, WS, SM	96.525	**91.069**	**94.155**

Conclusions

- Accuracy of ~98% was achieved by Naive Bayes classifier with all features used

- Average accuracy of ~87% was achieved by Random Forrest classifier when *Temperature* and *Light* features were not used

- Normalizing features improves the performance of SVM. Average accuracy of ~93% was achieved when *Temperature* and *Light* measurements were not used

Conclusions

- For measurements from one sensor only, Naïve Bayes method predicted occupancy the best
- *Seconds from Midnight* and *Week Status* turned out to be strong predictors of occupancy
- As it was expected, all algorithms showed better performance on Testing dataset 1, which had similar conditions to the training dataset

References:

[1] F. Alam et al. “Data Fusion and IoT for Smart Ubiquitous Environments: A Survey” IEEE Access, 2017

[2] Luis M. Candanedo, Véronique Feldheim, “Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models”, Energy and Buildings, Volume 112, 15 January 2016, Pages 28-39, ISSN 0378-7788

[3] Liu, J., et al., “Fuzzy logic controller for energy savings in a smart LED lighting system considering lighting comfort and daylight”, Energy and Buildings, 2016. 127: p. 95-104.