Developing a robust system for occupancy detection in the household

Ash Tyndall

This report is submitted as partial fulfilment
of the requirements for the Honours Programme of the
School of Computer Science and Software Engineering,
The University of Western Australia,
2015

Abstract

This is the abstract.

Keywords: keyword, keyword
CR Categories: category, category

Acknowledgements

These are the acknowledgements.

Contents

List of Tables

List of Figures

List of Listings

CHAPTER 1

Introduction

The proportion of elderly and mobility-impaired people is predicted to grow dramatically over the next century, leaving a large proportion of the population unable to care for themselves, and also reducing the number of human carers available [8]. With this issue looming, investments are being made in technologies that can provide the support these groups need to live independent of human assistance.

With recent advance in low cost embedded computing, such as the Arduino and Raspberry Pi, the ability to provide a set of interconnected sensors, actuators and interfaces to enable a low-cost 'smart home for the disabled' that takes advantage of the Internet of Things (IoT) is becoming increasingly achievable.

Sensing techniques to determine occupancy, the detection of the presence and number of people in an area, are of particular use to the elderly and disabled. Detection can be used to inform various devices that change state depending on the user's location, including the better regulation energy hungry devices to help reduce financial burden. Household climate control, which in some regions of Australia accounts for up to 40% of energy usage [5] is one area in which occupancy detection can reduce costs, as efficiency can be increased with annual energy savings of up to 25% found in some cases [7].

While many of the above solutions achieve excellent accuracies, in many cases they suffer from problems of installation logistics, difficult assembly, assumptions on user's technology ownership and component cost. In a smart home for the disabled, accuracy is important, but accessibility is paramount.

The goal of this research project is to devise an occupancy detection system that forms part of a larger 'smart home for the disabled', and intergrates into the IoT, that meets the following qualitative accessibility criteria;

- *Low Cost*: The set of components required should aim to minimise cost, as these devices are intended to be deployed in situations where the serviced user may be financially restricted.

- *Non-Invasive*: The sensors used in the system should gather as little information as necessary to achieve the detection goal; there are privacy concerns with the use of high-definition sensors.

- *Energy Efficient*: The system may be placed in a location where there is no access to mains power (e.g. roof), and the retrofitting of appropriate power can be difficult; the ability to survive for long periods on only battery power is advantageous.

- *Reliable*: The system should be able to operate without user intervention or frequent maintenance, and should be able to perform its occupancy detection goal with a high degree of accuracy.

To create a picture of what options there are in this sensing area, a literature review of the available sensor types and wireless sensor architectures is needed. From this list, proposed solutions will be compared against the aforementioned accessibility criteria to determine their suitability.

CHAPTER 2

Literature Review

To achieve the accessibility criteria, a wide variety of sensing approaches must be considered. It can be difficult to approach the board variety of sensor types in the field, so a structure must be developed through which to evaluate them. Teixeira, Dublon and Savvides [25] propose a 5-element human-sensing criteria which provides a structure through which we may define the broad quantitative requirements of different sensors.

These quantitative requirements can be used to exclude sensing options that clearly cannot meet the requirements before the more specific qualitative accessibility criteria will be considered for those remaining sensors.

The quantitative criteria elements are;

1. *Presence*: Is there any occupant present in the sensed area?
2. *Count*: How many occupants are there in the sensed area?
3. *Location*: Where are the occupants in the sensed area?
4. *Track*: Where do the occupants move in the sensed area? (local identification)
5. *Identity*: Who are the occupants in the sensed area? (global identification)

At a fundamental level, this research project requires a sensor system that provides both Presence and Count information. To assist with the reduction of privacy concerns, excluding systems that permit Identity will generally result in a less invasive system also. The presence of Location or Track are irrelevant to our project's goals, but overall, minimising these elements should in most cases help to maximise the energy efficiency of the system also.

Teixeira, Dublon and Savvides [25] also propose a measurable occupancy sensor taxonomy (see Figure 2.1 on the following page), which categorises different

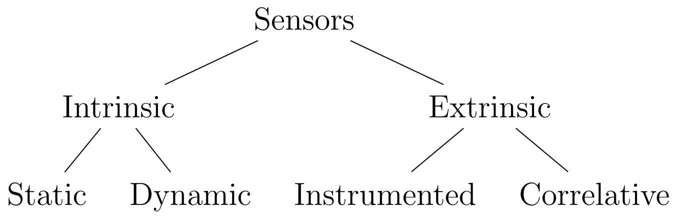


Figure 2.1: Taxonomy of occupancy sensors

sensing systems in terms of what information they use as a proxy for human-sensing. We use this taxonomy here as a structure through which we group and discuss different sensor types.

2.1 Intrinsic traits

Intrinsic traits are those which can be sensed that are a direct property of being a human occupant. Intrinsic traits are particularly useful, as in many situations they are guaranteed to be present if an occupant is present. However, they do have varying degrees of detectability and differentiation between occupants. Two main subcategories of these sensor types are static and dynamic traits.

2.1.1 Static traits

Static traits are physiologically derived, and are present with most (living) occupants. One key static trait that can be used for occupant sensing is that of thermal emissions. All human occupants emit distinctive thermal radiation in both resting and active states. The heat signatures of these emissions could potentially be measured with some apparatus, counted, and used to provide Presence and Count information to a sensor system, without providing Identity information.

Beltran, Erickson and Cerpa [7] propose Thermosense, a system that uses a type of thermal sensor known as an Infrared Array Sensor (IAR). This sensor is much like a camera, in that it has a field of view which is divided into “pixels”; in this case an 8×8 grid of detected temperatures. This sensor is mounted on an embedded device on the ceiling, along with a Passive Infrared Sensor (PIR), and uses a variety of classification algorithms to detect human heat signatures within the raw thermal and motion data it collects. Thermosense achieves Root Mean Squared Error ≈ 0.35 persons, meaning the standard deviation between Thermosense’s occupancy predictions and the actual occupancy number was ≈ 0.35.

Another static trait is that of CO_2 emissions, which, like thermal emissions, are emitted by human occupants in both resting and active states. By measuring the buildup of CO_2 within a given area, one can use a variety of mathematical models of human CO_2 production to determine the likely number of occupants present. Hailemariam et al. [14] trialled this as part of a sensor fusion within the context of an office environment, achieving a $\approx 94\%$ accuracy. Such a sensing system could provide both the Presence and Count information, and exclude the Identity information as required. However, a CO_2 based detection mechanism has serious drawbacks, discussed by Fisk, Faulkner and Sullivan [10]: The CO_2 feedback mechanism is very slow, taking hours of continuous occupancy to correctly identify the presence of people. In a residential environment, occupants are more likely to be moving between rooms than an office, so the system may have a more difficult time detecting in that situation. Similarly, such systems can be interfered with by other elements that control the CO_2 buildup in a space, like air conditioners, open windows, etc. This is also much more of a concern in a residential environment compared to the studied office space, as the average residence can have numerous such confounding factors that cannot easily be controlled for.

Visual identification can be, achieved through the use of video or still-image cameras and advanced image processing algorithms. Video can be used in occupancy detection in several different ways, achieving different levels of accuracy and requiring different configurations. The first use of video, POEM, proposed by Erickson, Achleitner and Cerpa [9] is the use of video as a "optical turnstile"; the video system detects potential occupants and the direction they are moving in at each entrance and exit to an area, and uses that information to extrapolate the number of occupants within the turnstiled area; this system has up to a 94% accuracy. However, the main issue with such a system applied to a residential environment is the system assumes that there will be wide enough "turnstile areas", corridors of a fairly large area that connect different sections of a building, to use as detection zones. While such corridors exist in office environments, they are less likely to exist in residential ones.

Another video sensor system is proposed by Serrano-Cuerda et al. [22], that uses ceiling-based cameras and advanced image processing algorithms to count the number of people in the captured area. This system achieves a specificity of $TP/(TP + FP) \approx 97\%$ and a sensitivity $TP/(TP + FN) \approx 96\%$ (TP = true positives, FP = false positives, FN = false negatives). Such a system could be successfully applied to the residential environment, as both it and the "optical turnstile" model provide Presence and Count information. However, these systems also allow Identity to be determined, and thus are perceived as privacy-invasive. This perception leads to adoption and acceptance issues, which work against the ideal system's goals.

2.1.2 Dynamic traits

Dynamic traits are usually products of human occupant activity, and thus can generally only be detected when a human occupant is physically active or in motion.

Ultrasonic systems, such as Doorjamb proposed by Hnat et al. [15], use clusters of such sensors above doorframes to detect the height and direction of potential occupants travelling between rooms. This acts as a turnstile based system, much like POEM [9], but augments this with an understanding of the model of the building to error correct for invalid and impossible movements brought about from sensing errors. This system provides an overall room-level tracking accuracy of 90%, however to achieve this accuracy, potential occupants are intended to be tracked using their heights, which has privacy implications. The system can also suffer from problems with error propagation, as there are possibilities of "phantom" occupants entering a room due to sensing errors.

Solely PIR based systems, like those used by Hailemariam et al. [14], involve the motion of the sensor being averaged over several different time intervals, and fed into a decision tree classifier. This PIR system alone produced a $\approx 98\%$ accuracy. However, such a system, due to only motion detection capabilities, can only provide Presence information, and is unable to provide Count information, nor detect motionless occupants.

2.2 Extrinsic traits

Extrinsic traits are those which are actually other environmental changes that are caused by or correlated with human occupant presence. These traits generally present a less accurate picture, or require the sensed occupants to be in some way "tagged", but they are generally also easier to sense in of themselves. The sensors in this category have been divided into two subcategories.

2.2.1 Instrumented traits

One extrinsic trait category is instrumented approaches; these require that detectable occupants carry with them some device that is detected as a proxy for the occupant themselves.

The most obvious of these approaches is a specially designed device. Li et al. [19] use RFID tags placed on building occupant's persons and a set of transmitters to triangulate the tags and place them within different thermal zones for the use of the HVAC system. For stationary occupants, there was a detection accuracy of $\approx$ 88%, and for occupants who were mobile, the accuracy was $\approx$ 62%. Such a system could be re-purposed for the residence, however, these systems raise issues in a residential environment as it requires occupants to be constantly carrying their sensors, which is less likely in such an environment. Additionally, the accuracy for this system is not necessarily high enough for a residential environment, where much smaller rooms are used.

To make extrinsic detection more reliable, Li, Calis and Becerik-Gerber [16] leverage a common consumer device; wifi enabled smart phones. They propose the *homeset* algorithm, which uses the phones to scan the visible wifi networks, and from that information estimate if the occupants are at home or out and about by "triangulating" their position from the visible wifi networks. This solution does not provide the fine-grained Presence data that we need, as it is only able to triangulate the phone's position very roughly with the wireless network detection information.

Balaji et al. [6] also leverage smart phones to determine occupancy, but in a more broad enterprise environment: Wireless association logs are analysed to determine which access points in a building a given occupant is connected to. If this access point falls within the radio range of their designated "personal space", they are considered to be occupying that personal space. This technique cannot be applied to a residential environment, as there are usually not multiple wireless hotspots.

Finally, Gupta, Intille and Larson [13] use specifically the GPS functions of the smartphone to perform optimisation on heating and cooling systems by calculating the "travel-to-home" time of occupants at all times and ensuring at every distance the house is minimally heated such that if the potential occupant were to travel home, the house would be at the correct temperature when they arrived. While this system does achieve similar potential air-conditioning energy savings, it is not room-level modular, and also presupposes an occupant whose primary energy costs are from incorrect heating when away from home, which isn't necessarily the case for this demographic.

2.2.2 Correlative traits

The second of these subcategories are correlative approaches. These approaches analyse data that is correlated with human occupant activity, but does not require

	Requires		Excludes	Irrelevant	
	Presence	Count	Identity	Location	Track
<u>Intrinsic</u>					
Static					
Thermal	✓	✓	✓	✓	
CO_2	✓	✓	✓		
Video	✓	✓	✗	✓	✓
Dynamic					
Ultrasonic	✓	✓	✗		✓
PIR	✓	✗	✓		
<u>Extrinsic</u>					
Instrumented					
RFID	✓[1]	✓	✓	✓	
WiFi assoc.[2]	✓[1]	✓	✗	✓	
WiFi triang.[2]	✓[1]	✓	✗		
GPS[2]	✓[1]	✗	✓	✓	
Correlative					
Electricity	✓[1]	✗	✓		

[1]Doesn't provide data at required level of accuracy for home use.
[2]Uses smartphone as detector.

Table 2.1: Comparison of different sensors and project requirements

a specific device to be present on each occupant that is tracked with the system.

The primary approach in this area is work done by Kleiminger et al. [17], which attempts to measure electricity consumption and use such data to determine Presence. Electricity data was measured at two different levels of granularity; the whole house level with a smart meter, and the consumption of specific appliances through smart plugs. This data was then processed by a variety of classifiers to achieve a classification accuracy of more than 80%. Such a system presents a low-cost solution to occupancy, however it is not sufficiently granular in either the detection of multiple occupants, or the detection of occupants in a specific room.

2.3 Analysis

From these various sensor options, there are a few candidates that provide the necessary quantitative criteria (Presence and Count); these are thermal, CO_2 , Video, Ultrasonic, RFID and WiFi association and triangulation based methods. All sensing options are compared on Table 2.1 on the previous page.

In the context of our four qualitative accessibility criteria, CO_2 sensing has several reliability drawbacks, the predominant ones being a large lag time to receive accurate occupancy information and interference from a variety of air conditioning sources which can modify the CO_2 concentration in the room in unexpected ways.

Video-based sensing methods suffer from invasiveness concerns, as they by design must have a constant video feed of all detected areas.

Ultrasonic methods suffer from reliability concerns when a user falls outside the prescribed height bounds of normal humans. Wheelchair bound occupants, a core demographic of our proposed sensing system, are not discussed in the Doorjamb paper. Their wheelchair may also interfere with height measurement results. Ultrasonic methods also provide weak Identity information through height detection.

RFID sensing also has several drawbacks; it is difficult value proposition to get residential occupants to carry RFID tags with them continuously. Another drawback is that the triangulation methods discussed are too unreliable to place occupants in specific rooms in many cases.

WiFi association is not granular enough for residential use, as the original enterprise use case presupposed a much larger area, as well as multiple wireless access points, neither of which a typical residential environment have.

WiFi triangulation is a good candidate for residential use, as there are most likely neighbouring wireless networks that can be used as virtual landmarks. However, it suffers from the same granularity problems as WiFi association, as these signals are not specific enough to pinpoint an occupant to a specific room.

For approaches presupposing smartphones being present on each occupant, it is more difficult to ensure that occupants are carrying their smartphones with them at all times in a residential environment. Another issue with smart phones is that they represent an expense that the target markets of the elderly and the disabled may not be able to afford.

Finally, we have thermal sensing. It provides both Presence and Count information, as it uses occupants' thermal signatures to determine the presence of people in a room. It does not however provide Identity information, as thermal signatures are not sufficiently unique with the technologies used to distinguished between occupants. Such a sensor system is presented as low-cost and energy efficient within Thermosense [7], is non-invasive by design and can reliably detect occupants with a very low root mean squared error. For our specific accessibility criteria, thermal sensing appears to be the best option available.

2.4 Thermal sensors

Our analysis (Subsection 2.3 on page 8) concluded that thermal sensors are the best candidates for this project. In this section we discuss the thermal sensing field in more detail.

A primary static/dynamic sensor fusion system in this field is the Thermosense system [7], a Passive Infrared Sensor (PIR) and Infrared Array Sensor (IAR) used to subdivide an area into an 8×8 grid of sections from which temperatures can be derived. This sensor system is attached to the roof on a small embedded controller which is responsible for collecting the data and transmitting it back to a larger computer via low powered wireless protocols.

The Thermosense system develops a thermal background map of the room using an Exponential Weighted Moving Average (EMWA) over a 15 minute time window (if no motion is detected). If the room remains occupied for a long period, a more complex scaling algorithm is used which considers the coldest points in the room empty, and averages them against the new background, then performs EMWA with a lower weighting.

This background map is used as a baseline to calculate standard deviations of each grid area, which are then used to determine several characteristics to be used as feature vectors for a variety of classification approaches. The determination of the feature vectors was subject to experimentation, since the differences at each grid element too susceptible to individual room conditions to be used as feature vectors. Instead, a set of three different features was designed; the number of temperature anomalies in the space, the number of groups of temperature anomalies, and the size of the largest anomaly in the space. These feature vectors were compared against three classification approaches; K-Nearest Neighbors, Linear Regression and an a feed-forward Artificial Neural Network of one hidden later and 5 perceptions. All three classifiers achieved a Root Mean Squared Error (RMSE) within 0.38 ± 0.04. This final classification is subject to a final averaging process over a 4 minute window to remove the presence of independent errors from the raw classification data.

The Thermosense approach presents the state of the art in the field of sensing with IAR technology. Using a similar IAR system along with those types of classification algorithms should yield useful sensing results which can be then integrated into the broader sensor system.

2.5 Research Gap

Throughout this review of the area of wireless occupancy sensors within the Internet of Things (IoT) it can be seen that there is a clear research gap within the area of occupancy. No group could be found who has assembled an occupancy sensor that optimises these ares of Low Cost, Non-Invasiveness, Energy Efficiency and Reliability into a architected software and hardware package that can be integrated like any other Thing into the IoT.

This is a key research area, because, as we have previously mentioned, the true "disruptive level of innovation"[4] the IoT provides can only be realised once a novel idea has been properly packaged as a Thing, rather than as a research curiosity. Packaging something as a Thing requires careful consideration of the best sensing systems, the best hardware to run those systems on, the best protocols to allow these Things to communicate, and the best device architecture to enable that communication. The state of the art in all these areas have been discussed throughout this literature review.

2.6 Conclusion

Several criteria were identified through which the spectrum of occupancy sensing could be examined; a quantitative criteria by Teixeira, Dublon and Savvides [25] to examine the different functionality offerings of sensor systems and a qualitative criteria derived from the aims of the project to examine how those sensors fit within the project's parameters.

Occupancy research performed with different sensor types was examined methodically through a set of taxonomic categories also originally proposed by Teixeira, Dublon and Savvides [25], but modified to better suit the specifics of occupancy sensors. These sensor types included Thermal, CO_2 , Video, Ultrasonic, Passive Infrared Sensor (PIR), RFID, various WiFi based methods, GPS and electricity consumption. Through an examination of these sensing systems quantitative and qualitative characteristics, it was determined that the Thermosense Infrared Array Sensor (IAR) system [7] was the most suitable to the project's aims.

A key part of enabling the "smart home for the disabled" is creating a set of Things that can improve quality of life for those people. We believe our proposed Thing has clearly demonstrated this potential.

CHAPTER 3

Prototype Design

As discussed in the Literature Review, using an Infrared Array Sensor (IAR) appear to be the most viable way to achieve the high-level goals of this project. Thermosense [7], the primary occupancy sensor in the IAR space, used the low-cost Panasonic Grid-EYE sensor for this task. This sensor, costing around $30USD, appears to be a prime candidate for use in this project, as it satisfied low-cost criteria, as well as being proven by Thermosense to be effective in this space. However, while still available for sale in the United States, we were unable to order the sensor for shipping to Australia due to export restrictions outside of our control. While such restrictions would be circumventable with sufficient effort, using a sensor with such restrictions in place goes against an implicit criteria of the parts used in the project being relatively easy to acquire.

This forced us to search for alternative sensors in the space that fulfill similar criteria but were more broadly available. The sensor we settled on was the Melexis MLX90620 (*Melexis*) [20], an IAR with similar overall qualities that differed in several important ways; it provides a 16×4 grid of thermal information, it has an overall narrower field of view and it sells for approximately $80USD. Like the Grid-EYE , the *Melexis* sensor communicates over the 2-wire I^2C bus, a low-level bi-directional communication bus widely used and supported in embedded systems.

In an idealized version of this occupancy system, much like Thermosense this system would include wireless networking and a very small form factor. However, due to time and resource constraints, the scope of this project has been limited to a minimum viable implementation. Appendix Chapter B on page 61 discusses in detail how the introduction of new open standards in the Wireless Personal Area Network space could be used in future systems to provide robust, decentralized networking of future occupancy sensors. This prototype architecture has been designed such that a clear path to the idea system architecture discussed therein is available.

Analysis Tier	Raspberry Pi B+
Preprocessing Tier	Arduino Uno R3
Sensing Tier	Melexis MLX90620 & PIR

Table 3.1: Hardware tiers

3.1 Hardware

As reliability and future extensibility are core concerns of the project, a three-tiered system is employed with regards to the hardware involved in the system (Table 3.1). At the bottom, the Sensing Tier, we have the raw sensor, the Melexis MLX90620 (*Melexis*), which communicate over I^2C . Connected to these devices via those respective protocols is the Preprocessing Tier, run an embedded system. The embedded device polls the data from these sensors, performs necessary calculations to turn raw information into suitable data, and communicates this via Serial over USB to the third tier. The third tier, the Analysis Tier, is run on a fully fledged computer. In our prototype, it captures and stores both video data, and the Temperature and Motion data it receives over Serial over USB.

While at a glance this system may seem overly complicated, it ensures that a sensible upgrade path to a more feature-rich sensing system is available. In the current prototype, the Analysis Tier merely stores captured data for offline analysis, in future prototypes this analysis can be done live and served to interested parties over a RESTful API. In the current prototype, the Analysis and Sensing Tiers are connected by Serial over USB, in future prototypes, this can be replaced by a wireless mesh network, with many Preprocessing/Sensing Tier nodes communicating with one Analysis Tier node.

Due to low cost and ease of use, the Arduino platform was selected as the host for the Preprocessing Tier, and thus the low-level I^2C interface for communication to the *Melexis*. Initially, this presented some challenges, as the *Melexis* recommends a power and communication voltage of 2.6V, while the Arduino is only able to output 3.3V and 5V as power, and 5V as communication. Due to this, it was not possible to directly connect the Arduino to the *Melexis*, and similarly due to the two-way nature of the I^2C 2-wire communication protocol, it was also not possible to simply lower the Arduino voltage using simple electrical techniques, as such techniques would interfere with two-way communication.

A solution was found in the form of a I^2C level-shifter, the Adafruit "4-channel I2C-safe Bi-directional Logic Level Converter" [1], which provided a cheap method to bi-directionally communicate between the two devices at their own preferred voltages. The layout of the circuit necessary to link the Arduino

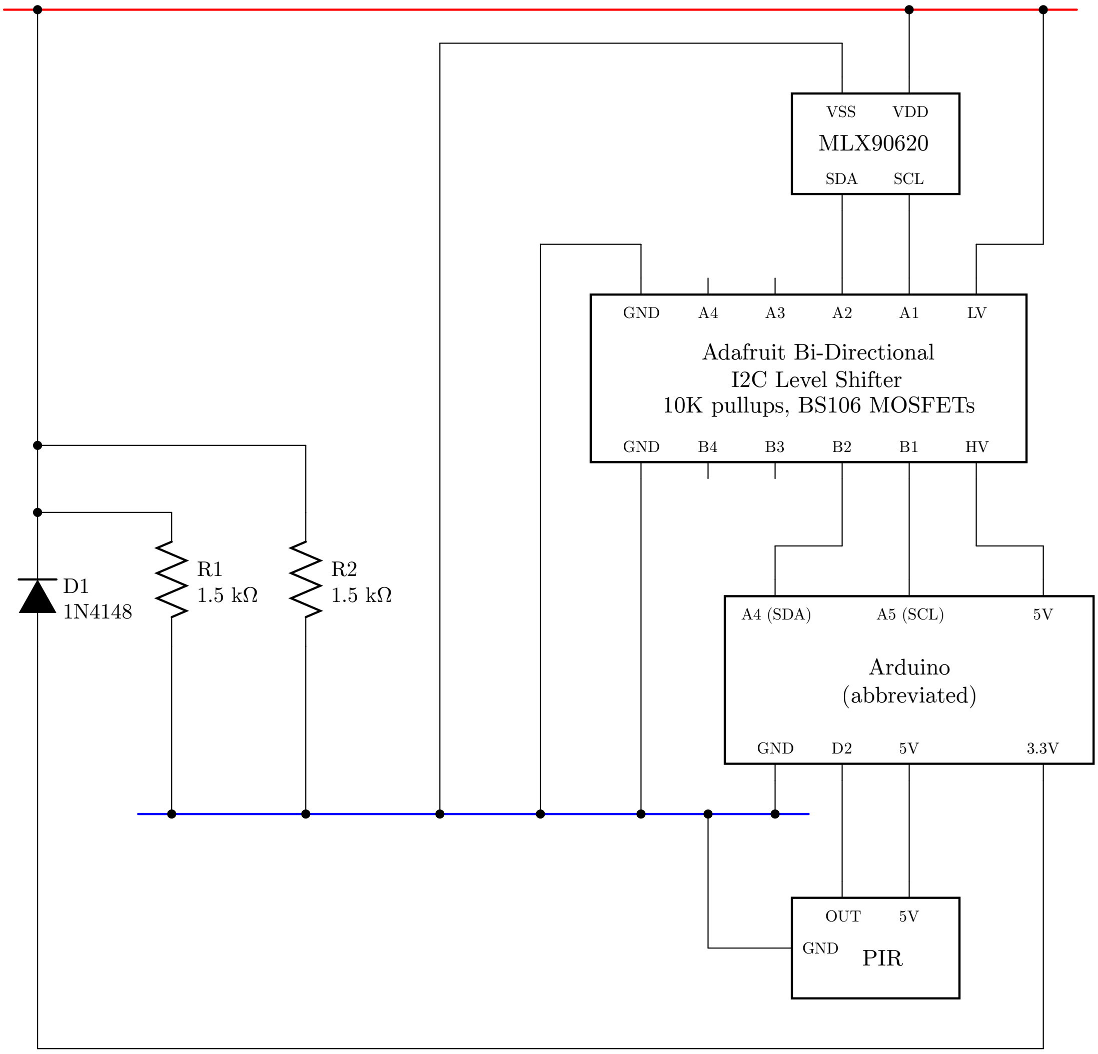


Figure 3.1: MLX90620, PIR and Arduino integration circuit

and the *Melexis* using this converter can be seen in Figure 3.1 on the previous page.

Additionally, as used in the Thermosense paper, a Passive Infrared Sensor (PIR) motion sensor [2] was also connected to the Arduino . This sensor, operating at 5V natively, did not require any complex circuitry to interface with the Arduino . It is connected to digital pin 2 on the Arduino , where it provides a rising signal in the event that motion is detected, which can be configured to cause an interrupt on the Arduino . In the configuration used in this project, the sensor's sensitivity was set to the highest value and the timeout for re-triggering was set to the lowest value (approximately 2.5 seconds). Additionally, the continuous re-triggering feature (whereby the sensor produces continuous rising and falling signals for the duration of motion) was disabled using the provided jumpers.

For the Analysis Tier, the Raspberry Pi B+ was chosen, as it is a powerful computer capable of running Linux available for an extraordinarily low price. The Arduino is connected to the Raspberry Pi over USB, which provides it both power and the capacity to transfer data. In turn, the Raspberry Pi is connected to a simple micro-USB rechargeable battery pack, which provides it with power, and subsequently the Arduino and sensors.

3.2 Software

At each layer of the described three-tier software architecture (pictured in greater detail in Figure 3.2 on page 17), there must exist software which governs the operation of that tier's processing concerns. Software in this project was written in two different languages.

At the Sensing Tier, it was not necessary for any software to be developed, as any software necessary came pre-installed and ready for use on the aforementioned sensors.

At the Preprocessing Tier, the Arduino, the default C++ derivative language was used, as careful management of memory usage and algorithmic complexity is required in such a resource-constrained environment, thus limiting choice in the area.

Finally, at Analysis Tier, a computer running fully-fledged Linux, choice of language becomes a possibility. In this instance, Python was settled on as the language of choice, as it is a quite high-level language with excellent library support for the functions required of the Analysis Tier, including serial interface, the use of the Raspberry Pi's built in camera, and image analysis. The 2.x branch of Python was chosen over the 3.x branch, despite its age, due a greater maturity in support for several key graphical interface libraries.

3.2.1 Pre-processing: `mlx90620_driver.ino`

On the Arduino, once large program was developed, termed `mlx90620_driver.ino`. This program's purpose was to take simple commands over serial to configure the Melexis MLX90620 (*Melexis*) and to report back the current temperature values and Passive Infrared Sensor (PIR) motion information at either a pre-set interval, or when requested.

To calculate the final temperature values that the *Melexis* offers, a complex initialization and computational process must be followed, which is specified in the sensor's datasheet [20]. This process involves initializing the sensor with values attained from a separate on-board I^2C EEPROM, then retrieving a variety of normalization and adjustment values, along with the raw sensor data, to compute the final temperature result.

The basic algorithm to perform this normalization was based upon the provided datasheet [20], as well as code by users "maxbot", "IIBaboomba", "nseidle" and others on the Arduino Forums [3] and was modified to operate with the newer Arduino "Wire" I^2C libraries released since the authors' posts. In pursuit of the project's aims to create a more approachable thermal sensor, the code was also restructured and rewritten to be both more readable, and to introduce a set of features to make the management of the sensor data easier for the user, and for the information to be more human readable.

Additionally, support for the PIR's motion data was added to the code, with the PIR configured to perform interrupts on one of the Arduino's digital pinsnd the code structured to take note of this information and to report it to the user in the "MOTION" section of the next packet.

The first of the features introduced was the human-readable format for serial transmission. This allows the user to both easily write code that can parse the serial to acquire the serial data, as well as examine the serial data directly with ease. When the Arduino first boots running the software, the output in Figure 3.3 on page 18 is output. This specifies several things that are useful to the user; the attached sensor ("DRIVER"), the build of the software ("BUILD") and the refresh rate of the sensor ("IRHZ"). Several different headers, such as "ACTIVE" and "INIT" specify the current millisecond time of the processor, thus indicating how long the execution of the initialization process took (33 milliseconds).

Once booted, the user is able to send several one-character commands to

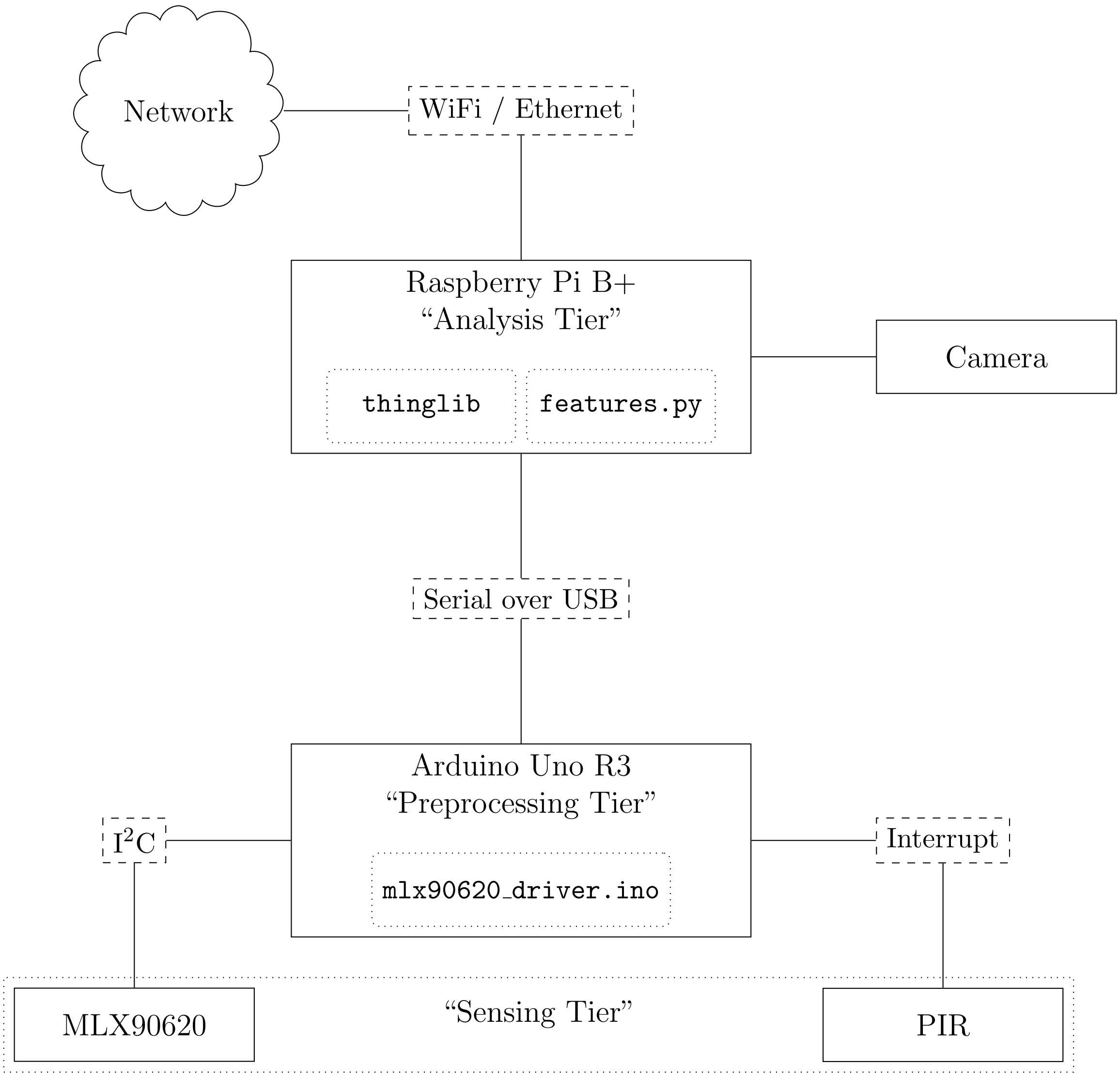


Figure 3.2: Prototype system architecture

```
INIT 0
INFO START
DRIVER MLX90620
BUILD Feb  1 2015 00:00:00
IRHZ 1
INFO STOP
ACTIVE 33

START 34
MOVEMENT 0
1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
STOP 97
```

Figure 3.3: Initialisation sequence and thermal packet

the sensor to configure operation, which are described in Table **??** on page ??. Depending on the sensor configuration, IR data may be periodically output automatically, or otherwise manually triggered. This IR data is produced in the packet format described in Figure 3.3. This is a simple, human readable format that includes the millisecond time of the processor at the start and end of the calculation, if the PIR has seen any motion for the duration of the calculation, and the 16x4 grid of calculated temperature values.

3.2.2 Analysis: `thinglib`

On the analysis tier a set of Python libraries and accompanying utility scripts were developed to interface with the Arduino, parse and interpret its data, and to provide data logging and visualization capabilities. Most of this functionality was split into a reusable and versatile Python module called `thinglib`.

`thinglib` provides 4 main feature sets across 3 files; the `Manager` series of classes, the `Visualizer` class, the `Features` class and the `pxdisplay` module.

3.2.2.1 `Manager` classes

The Manager series of classes are the direct interface between the Arduino and the Python classes. They implement a multi-threaded serial data collection and parsing system which converts the raw serial output of the connected Arduino into a series of Python data structures that represent the collected temperature and motion data of each captured frame. Several different versions of the `Manager` class exist to perform slightly different functions. When initializing these classes the sample rate of the *Melexis* can be configured, and it will be sent through to the Arduino for updating.

`BaseManager` is responsible for the implementation of the core serial parsing functions. It also provides a threaded interface through which the *Melexis*'s continuous stream of data can be subscribed to by other threads. The primary API, the `subscribe_` series of functions, return a thread-safe queue structure, through which thermal packets can be received by various other threads when they become available.

`Manager`, the primary class, provides access the *Melexis*'s data at configurable intervals. When initializing this class, you may specific 0.5, 1, 2, 4 or 8Hz, and the class will configure the Arduino to both set the *Melexis* to this sample rate, and to automatically write this data to the serial buffer whenever it is available. This serial interface is multi-threaded, as at higher serial baud rates if data was not polled continuously enough the internal serial buffer would fill and some data would be discarded. By ensuring this process cannot be blocked by other parts of the running program this problem is mostly eliminated.

`OnDemandManager` operates in a similar way to `Manager`, however instead of using a non-blocking threaded approach, the user's scripts may request thermal/motion data from the class, and it will poll the Arduino for information and block until this information is parsed and returned.

Finally, `ManagerPlaybackEmulator` is a simple class which can take a previously created thermal recording from a file, and emulate the `Manager` class by providing access to thread-safe queues which return this data at the specified Hz rate. This class can be used as a means to playback thermal recordings with the same visualization functions.

3.2.2.2 `pxdisplay` functions

The `pxdisplay` module is a set of functions that utilize the `pygame` library to create a simple live-updating window containing a thermal map representation of the thermal data. One can generate any number of `pxdisplay` objects, which leverage the `multithreading` library and `multithreading.Queue` to allow thermal data to be sent to the display.

The class also provides a set of functions to set a "hotest" and "coldest" temperature and have RGB colors assigned from red to blue for each temperature that falls between those two extremes.

3.2.2.3 `Visualizer` class

The `Visualizer` class is the natural compliment to the `Manager` series of classes. The functions contained within can usually be provided with a Queue object (generated by a `Manager` class) and can perform a variety of visualization and storage functions.

From the recording side, the `Visualizer` class can "record" a thermal capture by saving the motion and thermal information to a simple `.tcap` file, which stores the sample rate, timings, thermal and motion data from a capture in a very straightforward format. The class can also read these files back into the data structures `Visualizer` uses internally to store data. If `Visualizer` is running on a Raspberry Pi, it can also leverage the `picamera` library an the `OnDemandManager` class to synchronously capture both visual and thermal data for ground truth purposes.

From the visualization side, `Visualizer` can leverage the `pxdisplay` module to create thermal maps that can update in real-time based on the thermal data provided by a `Manager` class. The class can also generate both images and movie files from thermal recordings using the PIL and ffmpeg libraries respectively.

3.2.2.4 `Features` class

In Thermosense [7], an algorithm was demonstrated that allowed the separation of "background" information from "active" pixels, and from that information, the extraction of the features necessary for a classifier to correctly determine the number of people in an 8×8 thermal image. This algorithm involved calculating the average and standard deviations of each pixel while it is guaranteed that the image would be empty, and then when motion is detected, considering any pixel "active" that reaches a value more than 3 standard deviations above the pixel when there was no motion.

From these "active" pixels, it was established that a set of three feature vectors were all that were required to correctly classify the number of people in the thermal image. These feature vectors were;

1. **Number of active pixels**: The total number of pixels that are considered "active" in a given frame

2. **Number of connected components**: If each active pixel is represented as an node in an undirected graph where adjacent active pixels are connected, how many connected components does this graph have?

3. **Size of largest connected component**: The number of active pixels contained within the largest connected component

In accordance with the pseudo-code outlined in the Thermosense paper, the algorithm described in Listing 3.1 on the following page was created to extract these figures. The portion of this code dealing with scaling the thermal background for rooms without motion was not implemented, as in all experiments tested, there exists a significant interval of time during which the no motion is guarenteed and the thermal background can be generated. The `networkx` library was used to generate the connected components information.

```python
import networkx, itertools

nomotion_wgt = 0.01
n_rows = 4
n_cols = 16
background = first_frame
means = first_frame
stds = [ [0]*16 ]*4
stds_post = [ [None]*16 ]*4

def create_features(new_frame, is_motion):
  active = []
  g = networkx.Graph()

  for i, j in itertools.product( range(n_rows), range(n_cols) ):
    prev = background[i][j]
    cur = new_frame[i][j]
    cur_mean = means[i][j]
    cur_std = stds[i][j]

    if not is_motion:
      background[i][j] = nomotion_wgt * cur + (1 - nomotion_wgt) * prev
      means[i][j]      = cur_mean + (cur - cur_mean) / n
      stds[i][j]       = cur_std + (cur - cur_mean) * (cur - means[i][j])
      stds_post[i][j]  = math.sqrt(stds[i][j] / (n-1))

    if (cur - background[i][j]) > (3 * stds_post[i][j]):
      active.append((i,j))
      g.add_node((i,j))

      # Add edges for nodes that have already been computed as active
      for ix, jx in [(-1, -1), (-1, 0), (-1, 1), (0, -1)]:
        if (i+ix, j+jx) in active:
          g.add_edge((i,j), (i+ix,j+jx))

  comps = list(networkx.connected_components(g))
  num_active = len(active)
  num_connected = len(comps)
  size_connected = max(len(c) for c in comps) if len(comps) > 0 else None

  return (num_active, num_connected, size_connected)
```

Listing 3.1: Core feature extraction code

3.3 Sensor Properties

In order to best utilize the Melexis MLX90620 (*Melexis*), we must first understand the properties it exhibits, and their potential affects on our ability to perform person related measurements. These properties can be broadly separated into three different categories; bias, noise and sensitivity. A broad range of data was collected with the sensor in a horizontal orientation using various sources of heat and cold to determine these properties. This experimental setup is described in Figure 3.4 on the next page.

3.3.1 Bias

When receiving no infrared radiation, the sensor should indicate a near-zero temperature. If in such conditions it does not, that indicates that the sensor has some level of bias in its measurement values. We attempted to investigate this bias by performing thermal captures of the night sky. While this does not completely remove the infrared radiation, it does remove a significant proportion of it.

In Table 3.2 on page 25 the thermal sensor was exposed to the night sky at a capture rate of 1Hz for 4 minutes, with the sensing results combined to create a set of means and standard deviations to indicate the pixels at "rest". The average temperature detected was 11.78°C, with the standard deviation remaining less than 0.51°C over the entire exposure period. The resultant thermal map shows that pixels centered around the four "primary" pixels in the center maintain a similar temperature around 9°C, with temperatures beginning to deviate as they became further from the center.

The most likely cause of this bias is related to the physical structure of the sensor. The *Melexis* is a rectangular sensor which has been placed inside a circular tube. Due to this physical arrangement, the sides of this rectangular sensor will be significantly closer to these edges than the center. If these sides are at an ambient temperature higher than the measurement data (as they were in this case) thermal radiation from the sensor package itself could provide significant enough to cause the edges to appear warmer than the observed area of the sky. Such issues with temperature could be controlled for using a device that cools the sensor package to below that of the ambient temperature being measured, however, this is not a concern in this project, as the method of calculating a thermal background will compensate for any such bias as long as it remains constant.

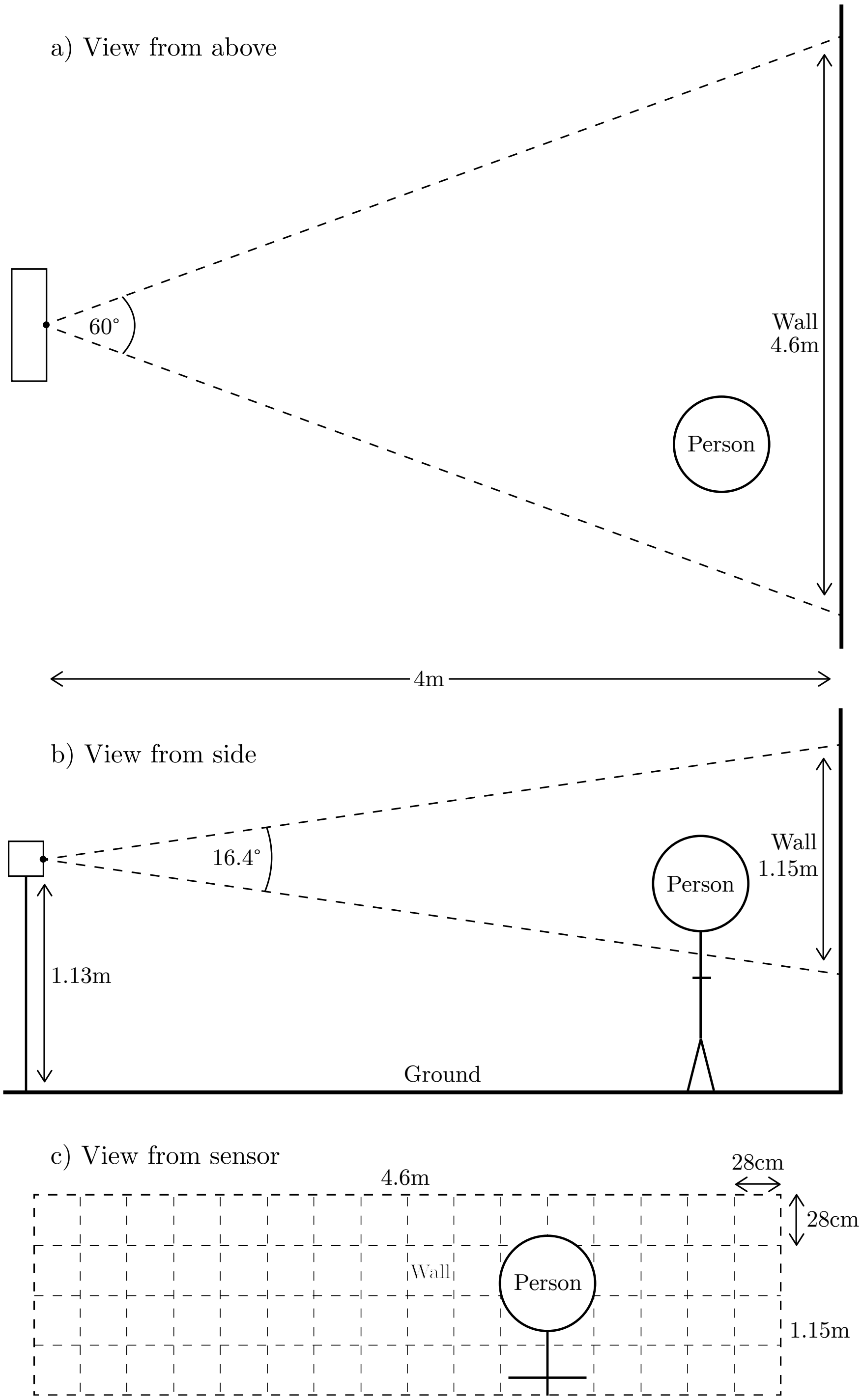


Figure 3.4: Experiment setup to determine sensor properties

14.95 **0.51**	**14.33** **0.27**	**12.34** **0.27**	**8.77** **0.33**	**8.15** **0.31**	**10.84** **0.38**	**9.02** **0.26**	**7.79** **0.37**	**6.67** **0.27**	**9.63** **0.29**	**9.29** **0.26**	**8.24** **0.27**	**9.84** **0.25**	**14.28** **0.33**	**14.92** **0.3**	**13.16** **0.25**
14.54 **0.34**	**15.62** **0.31**	**12.73** **0.23**	**11.51** **0.27**	**11.79** **0.26**	**11.47** **0.27**	**11.43** **0.29**	**9.02** **0.35**	**8.57** **0.23**	**11.15** **0.23**	**10.64** **0.22**	**10.3** **0.24**	**12.09** **0.22**	**14.49** **0.26**	**14.88** **0.31**	**14.71** **0.36**
18.25 **0.45**	**16.62** **0.31**	**14.15** **0.24**	**11.97** **0.34**	**13.11** **0.3**	**12.64** **0.22**	**10.66** **0.23**	**9.15** **0.24**	**9.58** **0.28**	**11.95** **0.28**	**11.22** **0.24**	**11.52** **0.36**	**11.11** **0.23**	**12.59** **0.25**	**14.44** **0.31**	**13.35** **0.28**
16.02 **0.28**	**16.81** **0.36**	**15.0** **0.25**	**11.53** **0.28**	**10.18** **0.29**	**12.2** **0.25**	**11.78** **0.29**	**8.36** **0.31**	**8.15** **0.33**	**10.36** **0.32**	**10.74** **0.31**	**8.25** **0.36**	**9.99** **0.35**	**12.42** **0.38**	**11.39** **0.4**	**11.06** **0.34**

Table 3.2: Mean and standard deviations for each pixel at rest

3.3.2 Noise

One of the features of the *Melexis* is the ability to sample the thermal data and a variety of sample rates between 0.5Hz and 512Hz. However, it was noted in early experimentation that a higher sample rate resulted in an increase in the noise contained within the resultant images. As our experiments focus on separating objects of interest from a thermal background, it is important to determine the maximum level of noise tolerable before our algorithms are unable to separate the background from the objects of interest.

Figure 3.5 on the next page plots one of the central pixels of the sensor in a scenario where it is merely viewing a background (shown in green), and when it is viewing a person (shown in red), at the 5 different sample rates achievable with the current hardware. We can see in these plots that the data becomes significantly more noisy as the sample rate increases, and we can also determine that the sensor uses a form of data smoothing at lower sample rates, as the variance in data increases with sample rate.

If the sample rate were to increase, it is likely that the ability for the sensing system to disambiguate between objects of interest and the background would diminish. However, in the current project, even the slowest sampling rate of 0.5Hz is sufficient, as occupancy estimations at a sub-second level present little additional value and would require significant reforms in the efficiency of the software used.

3.3.3 Sensitivity

The *Melexis* is a sensor composed of 64 independent non-contact digital thermopiles, which measure infrared radiation to determine the temperature of objects. While they are bundled in one package, Figure 3.6 on page 28 shows that they are in fact wholly independent sensors placed in a grid structure. This has important effects on the properties of the data that the *Melexis* produces.

Figure 3.7 on page 29 shows a graph of the temperatures of the top row of 16 pixels of the *Melexis* as a hot object is moved from left to right at an approximately similar speed. One of the most interesting phenomena in this graph is the apparent extreme variability of the detected temperature of the object as it moves "between" two different pixels; there is a noticeable drop in the objects detected temperature. Further analysis of each of the pixel's lines on the graph shows each pixel exhibiting a bell-curve like structure, with the detected temperature increasing from the baseline and peaking as the object enters the center of the pixel, and the detected temperature similarly decreasing

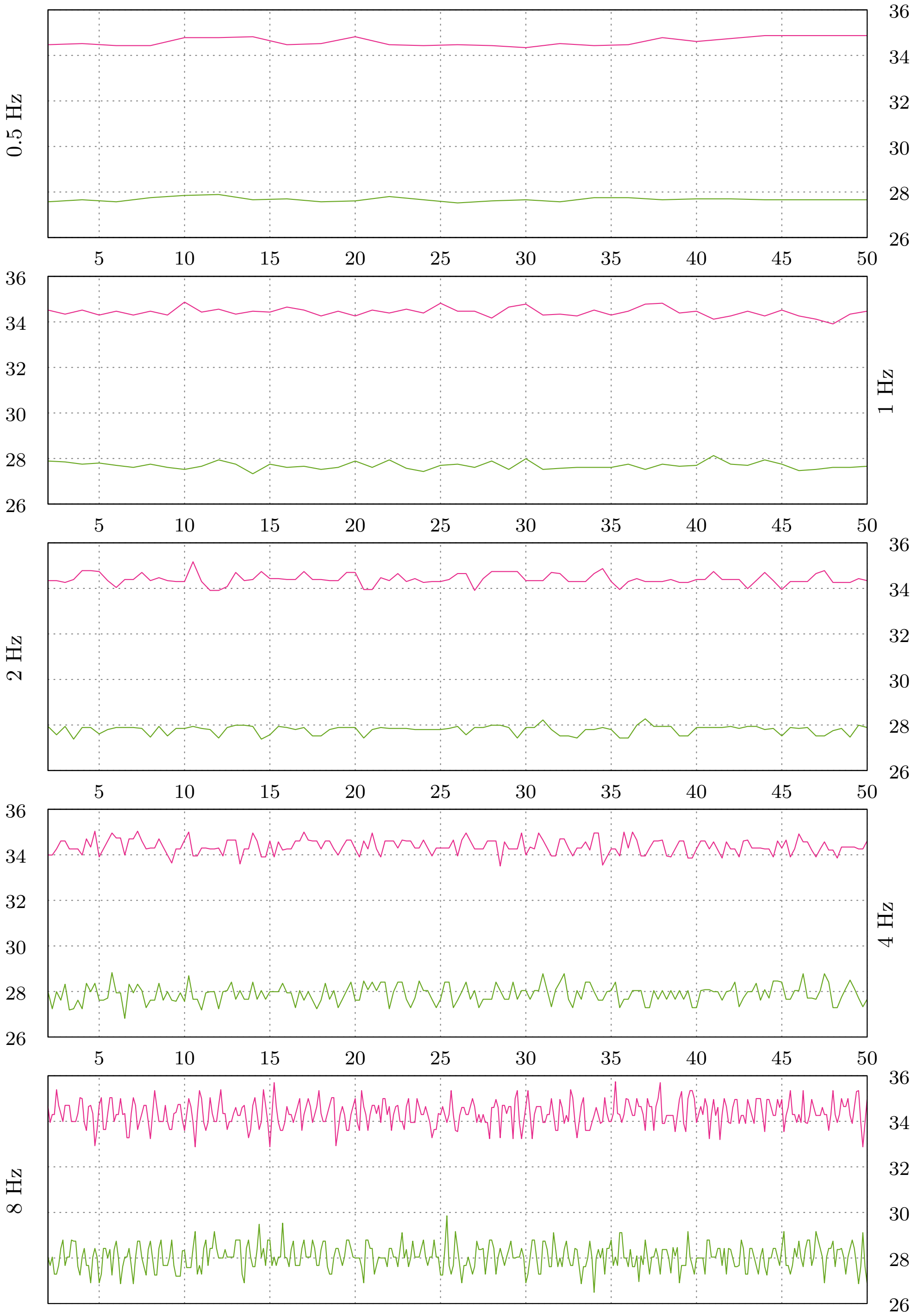


Figure 3.5: Comparison of noise levels at the *Melexis*' various sampling speeds

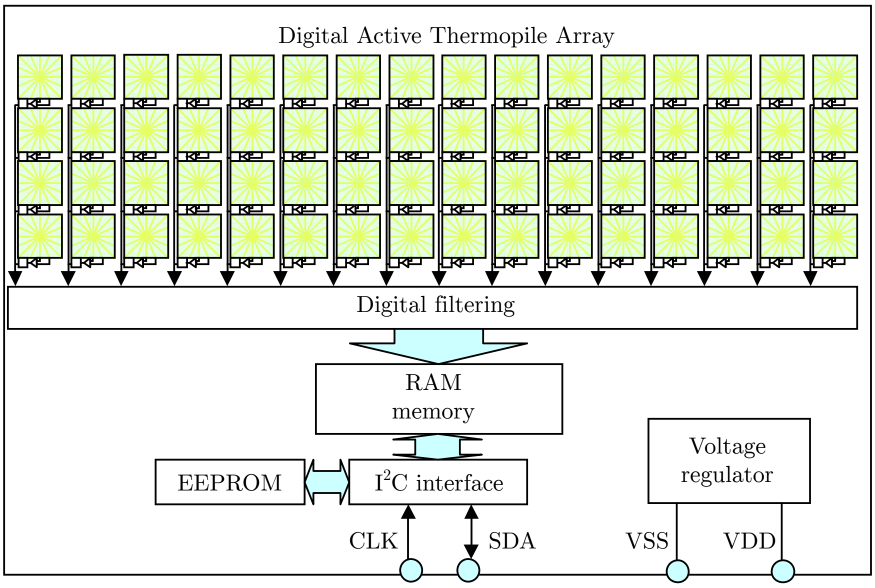


Figure 3.6: Block diagram for the *Melexis* taken from datasheet [20]

as the object leaves the center.

This phenomenon has several possible causes. One likely explanation is that each individual pixel detects objects radiating at less favorable angles of incidence to be colder than they actually are: As the object enters a pixel's effective field of view, it will radiate into the pixel at an angle that is at the edge of the pixel's ability to sense, with this angle slowing decreasing until the hot object is directly radiating into the pixel's sensor, causing a peak in the temperature reading. As the object leaves the individual elements field of view, the same happens in reverse.

While interesting, this phenomenon has little consequence to the effectiveness of the techniques used, as in experimental conditions the sensor will not be sufficiently distant that humans could be detected as single pixels. However, this phenomenon could be leveraged in future work to perform sub-pixel localization, discussed later on.

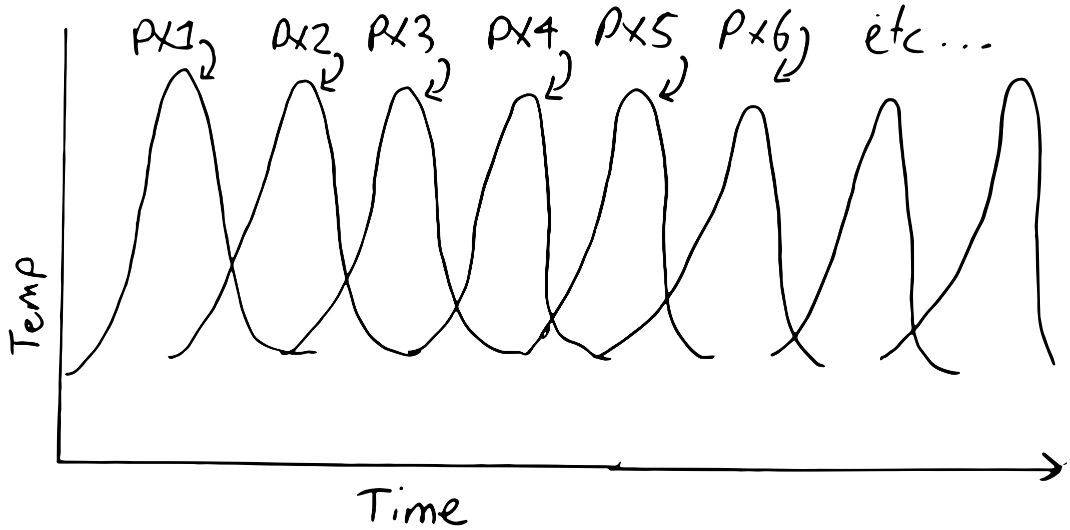


Figure 3.7: Different *Melexis* pixel temperature values as hot object moves across row

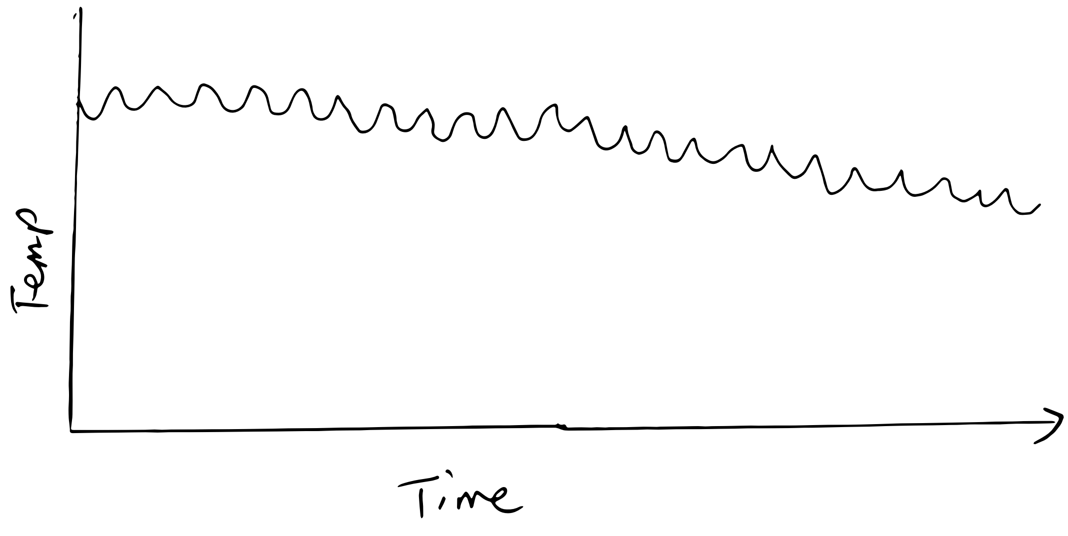


Figure 3.8: Variation in temperature detected for hot object at 1Hz sampling ration

CHAPTER 4

Methods

With a minimum viable prototype established, it now becomes possible to devise experimental scenarios to test against the project's goals. The project's adherence to the goals of Low-Cost and Non-Invasiveness have been evaluated previously, so in this section we will focus on the project's adherence to Reliability and Energy Efficiency goals.

4.1 Reliability Testing

With the prototype, it is now possible to utilize the prototype to gather both thermal and visual data in a synchronized format. This data can be collected and used to determine the effectiveness of the human counting algorithms used. Due to the prototype's technical similarly to Thermosense [7], a similar set of experimental conditions will be used, with a comparison against Thermosense being used as a benchmark. To this end, several experiments were devised, each of which had its data gathered and processed in accordance with the same general process, outlined in Figure 4.1 on the following page.

4.1.1 Data gathering

As the camera and the Arduino are directly plugged into the Raspberry Pi, all data capture is performed on-board through SSH, with the data being then copied of the Pi for later processing. To perform this capture, the main script used is `capture_pi_synced.py`.

`capture_pi_synced.py` takes two parameters on the command line; the name of the capture output, and the number of seconds to capture. By default, it always captures at 2Hz. The script initializes the `picamera` library, then passes a reference to it to the `capture_synced` function within the `Visualizer` class.

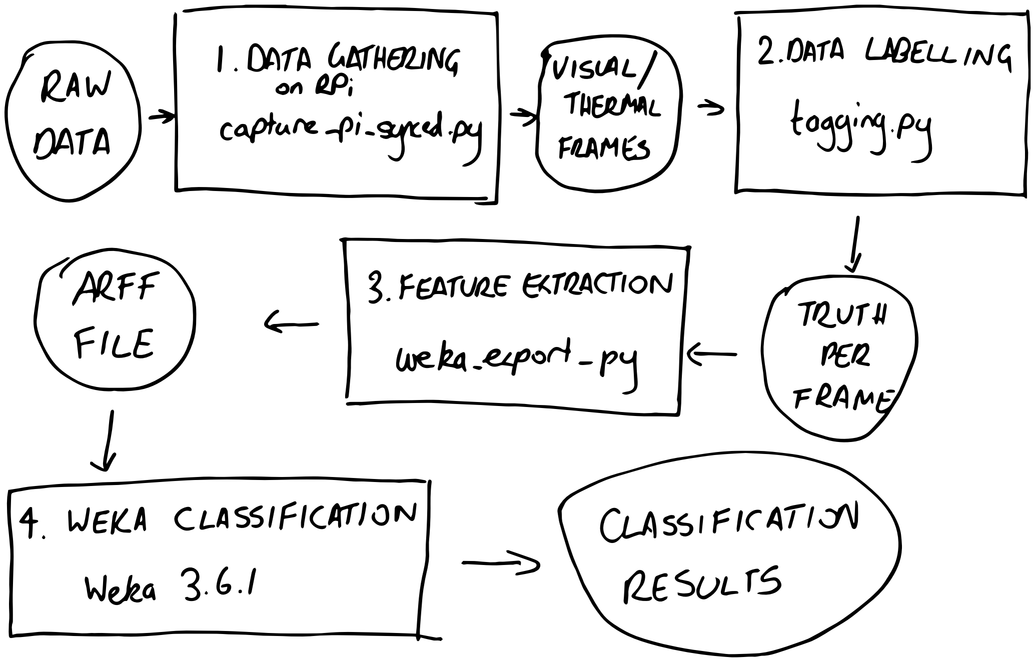


Figure 4.1: Flowchart of processing

The class will then handle the sending of commands to the Arduino to capture data in concert with taking still frames with the Raspberry Pi's camera.

When the script runs, it creates a folder with the name specified, storing inside a file named `output_thermal.hcap` containing the thermal capture, and a sequence of files with the format `video-%09d.jpg`, corresponding to each visual capture frame.

4.1.2 Data labeling

Once this data capture is complete, the data is copied to a more powerful computer for labeling. The utility `tagging.py` is used for this stage. This script is passed the path to the capture directory, and the number of frames at the beginning of the capture that are guaranteed to contain no motion. This utility will display frame by frame each visual and thermal capture together, as well as the computed feature vectors (based on a background map created from the first n frames without motion).

The user is then required to press one of the number keys on their keyboard to indicate the number of people present in this frame. This number will be recorded in a file called `truth` in the capture directory. The next frame will then be displayed, and the process continues. This utility enables the quick input of the ground truth of each capture, making the process more efficient.

4.1.3 Feature extraction and data conversion

Once the ground truth data is available, it is now possible to utilize the data to perform various classification tests. For this, we use version 3.7.12 of the open-source Weka toolkit [26], which provides easy access to a variety of machine learning algorithms and the tools necessary to analyze their effectiveness.

To enable the use of Weka, we export the ground truth and extracted features to a Comma Seperated Value (CSV) file for processing. `weka_export.py` takes two parameters, a comma-separated list of different experiment directories to pull ground truth and feature data from, and the number of frames at the beginning of each capture that can be considered as "motionless." With this information, a CSV-file file is generated on which the heading indicating the attribute names is added for Weka to recognize.

4.1.4 Running Weka Tests

Once the CSV file is generated, it is then possible to open the file in Weka for processing. Weka provides a variety of algorithms, but we choose a specific subset of algorithms based on those present in the Thermosense paper [7], as well others that we believe adequately represent the different approaches to classification.

4.1.4.1 Neural Networks

An artificial neural network (ANN) uses neurons as a model for machine learning. A number of input neurons, in this case connected to the feature vectors, is fed into a network of neurons (the "hidden layer"), each of which has an activation function which determines what set of inputs will make it fire. This network then connects to a number of output neurons which can be queried to determine the network's predicted result. In the nominal result case, there one neuron for each possible class, and in the numeric result case, there is one neuron without an activation function that outputs the raw numerical estimate. Neural networks can approximate functions of nearly any complexity with sufficient neurons in the correct topology, and are a quite common classification technique.

Thermosense uses a neural network with a hidden layer of five neurons, with a sigmoid activation function for the hidden layer and a linear activation function for the output layer. They test only the one, two and three person cases, relying on their Passive Infrared Sensor (PIR) to detect the zero person case. They use 70% of their data for training the neural net, 15% for testing the net and the final 15% for validating their results. Thermosense conducts tests interpreting the number of people as a numeric attribute.

We use Weka's "MultilayerPerceptron" neural network, which creates a hidden layer of $(attributes + classes)/2$ (three) by default, however we manually reconfigure this to be one hidden layer of five neurons, like Thermosense. It uses a sigmoid activation function for all neurons, except in the case that a numerical answer is to be predicted, in which case like Thermosense, it uses a linear activation function for the output layer. As is standard, for validation we use a 10-fold cross-validation for our nominal approach, and attempt to replicate Thermosense's configuration as closely as possible for the numeric result.

4.1.4.2 k-nearest Neighbors

A k-nearest Neighbors (KNN) approach uses the topology of the training data as a means to classify future data. For each data point that requires classification, a majority vote of its k nearest neighbors in the training data determines which class it belongs to. KNN is one of the simplest machine learning algorithms, and due to its classification method, is highly sensitive to classes that overlap.

Thermosense uses 5-nearest Neighbors with the Euclidean distance between points. For determining the class label, higher weightings are given to training points inversely to their distance from the point being classified. Thermosense appears to use a nominal classification for their KNN.

We use Weka's "iBk" function to perform a KNN calculation, configuring `distanceWeighting` to be "Weight by 1-distance" and `KNN` to be 5, to make the classification as similar in function to the Thermosense approach as is possible. Thermosense does not specify what validation technique they used, so we elected to use a standard 10-fold cross-validation.

4.1.4.3 Linear Regression

A Linear Regression approach attempts to construct a linear equation to describe the relationship between a dependent variable (in this case, the number of people in the space), and a number of other indicator variables (in this case, the three feature vectors). Generally, the equation takes the form $y = m_1x_1 + ... + m_nx_n + c$, where each of the feature vectors is multiplied by a weight, and then a final number is added to provide the final prediction.

Thermosense uses a Linear Regression model of $y = \beta_A A + \beta_S S + \beta$, whereby A is the number of active pixels, S is the size of the largest connected component, and the β values represent the corresponding coefficients. They opt to exclude the third feature, the number of connected components, as their testing indicates that excluding it minimizes the Root Mean Squared Error (RMSE) further.

We use Weka's "LinearRegression" function, exclude the `numconnected` attribute from the feature vector list, to attempt to match this approach.

4.1.4.4 Naive Bayes

A Naive Bayes approaches uses a simple application of Bayes' probability theorem to construct a probability of a given value belonging to a given class taking into account what is already known about the distribution of each of the classes in the data set, and the classification of those points that surround the point needing classification. One of the disadvantages of the Naive Bayes approach (the source of its naivety) is that it assumes independence between each of the variables used for classification.

In our data, the assumption of independence of variables is not correct, as each of the features are slightly different representations of the same data. However, due to Naive Bayes' ubiquity and simplicity, it can be illuminating to see how well a very common but poorly suited classifier fares with our data set. Within Weka, we use the "NaiveBayes" function, which has little by way of configuration, thus is left in its default state.

4.1.4.5 Support Vector Machines

Support Vector Machines (SVM) attempt to classify data by trying to find a plane that best separates two classes in a higher dimensional space. They do this by determining "support vectors," which are those data points that lie on the "edge" of the separation between classes, and then finding the plane that maximizes the margin between the two classes. We elected to test an SVM-based approach to determine if our data set is particularly suited to classification by SVMs.

For our purposes, we use Weka's "SMO" function, which implements the Sequential Minimal Optimization algorithm, an efficient and recent method of training SVMs. For datasets with more than two classes (such as ours), the "one vs. one" method is used, whereby an SVM is created for each pair of classes, and then a method of majority voting is used to determine which class is the ultimately correct one.

4.1.4.6 Decision Trees

A Decision Tree based approach uses the concept of a decision tree to create effectively a list of logical conditions which when met cause a data point to be classified as a specific class. Decision Tree classifiers generally use a partitioning approach whereby they split the data using a specific metric to maximize the tree's effectiveness. The advantages of Decision Trees are that they are considered to be "white boxes," specifically meaning that the result that they generate is human readable. This is useful, as in addition to the classifier providing its prediction of which class suits the data best, the tree can also be inspected to determine if the decisions it has extrapolated appear to be sensible, and even tweaked by humans if necessary.

One quite common algorithm for generating decision trees is C4.5, which is implemented by the "J48" function in Weka. C4.5 uses a measure of information gain, a concept rooted in information theory and entropy, to determine when to create splits in the tree. There are few configurable parameters for this approach, and for those we use the Weka defaults.

4.1.4.7 KStar

The KStar (K*) algorithm presents a change to the normal k-nearest Neighbors algorithm, in which the distance used to compare similar points is not the Euclidean distance, but rather an entropic distance. This has several positive effects; it makes the algorithm more robust to missing values, and also it makes the classifier able to output a numeric result in addition to or instead of a classification into a nominal class.

We have decided to use K* as one of our classification algorithms as it presents an interesting and different approach, and also allows the investigation of KNN-like techniques in the numeric area. K* is present in Weka as "KStar," and we will opt to use it in its default state.

4.1.4.8 0-R

0-R is our final classification algorithm. 0-R is a simple classifier that on nominal prediction will classify all new data as belonging to the category that was most common in the training data, and on numeric prediction will classify all new data as being the mean of all test data. A 0-R classifier, clearly, is not a serious classification technique, however it is useful in establishing a baseline from which to compare all other classification results.

In Weka, the 0-R classifier is known as "ZeroR" and accepts no parameters.

4.1.4.9 Excluding Zero

TODO: Talking about how there are two sets of data, that which includes the zero data, and that which excludes it and relies on the PIR to confirm zero. Pros and cons of these approaches.

Type	**Attribute**	**Weka Class & Parameters**
Neural Network (ANN)	Nominal, Numeric	`weka.classifiers.functions.MultilayerPerceptron` `-L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5`

k-nearest Neighbors (KNN)	Nominal, Numeric	`weka.classifiers.lazy.IBk`
		`-K 5 -W 0 -F`
		`-A "weka.core.neighboursearch.LinearNNSearch -A`
		`\"weka.core.EuclideanDistance -R first-last\""`

Naive Bayes	Nominal	`weka.classifiers.bayes.NaiveBayes`

Support Vector Machine (SVM)	Nominal	`weka.classifiers.functions.SMO`
		`-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1`
		`-K "weka.classifiers.functions.supportVector.PolyKernel`
		`-C 250007 -E 1.0"`

Decision Tree	Nominal	`weka.classifiers.trees.J48` `-C 0.25 -M 2`
Entropy Distance	Nominal, Numeric	`weka.classifiers.lazy.KStar` `-B 20 -M a`
Linear Regression	Numeric	`weka.classifiers.functions.LinearRegression` `-S 0 -R 1.0E-8`
0-R	Nominal, Numeric	`weka.classifiers.rules.ZeroR`

Table 4.1: Weka parameters used for classifications

For those tests that are "nominal," the `npeople` attribute was interpreted as nominal using the "NumericToNominal" filter, which creates a class for each value deleted in `npeople`'s columns. For those tests that are "numeric," `npeople` is left unchanged, as by default all CSV import attributes are interpreted as such. For all tests where not specifically instructed, we use 10-fold cross-validation to validate our results.

As the data we are using is based on real experiments, the number of frames which are classified as each class may be unbalanced, which could cause the classification results to be affected. To that end, for each classification technique, we both classify the data in its raw, unbalanced form, and we also uniformly resample the `npeople` parameter using `weka.filters.supervised.instance.Resample -B 1.0 -S 1 -Z 100.0` in the pre-processing stage.

To help maximize the efficiency of the classification task, we use the Weka Knowledge Flow constructor to generate an encompassing flow that accepts an input CSV file of the raw data, and performs all resampling, numeric and nominal classification, returning a text file with the results of each of the different classification techniques run. The knowledge flow's struture can be seen in . To enable maximum efficiency, the input and output elements of this flow are set to the environmental variables `UnifiedFlow.InputCSV` and `UnifiedFlow.OutputCSV`, a Jython script, `run_flow.py`, then sets those environmental variables to input and output file names, then calls the flow using Weka's Java API. After this is complete, the script then runs a series of regexes on the output text data to generate summary spreadsheets with the relevant values.

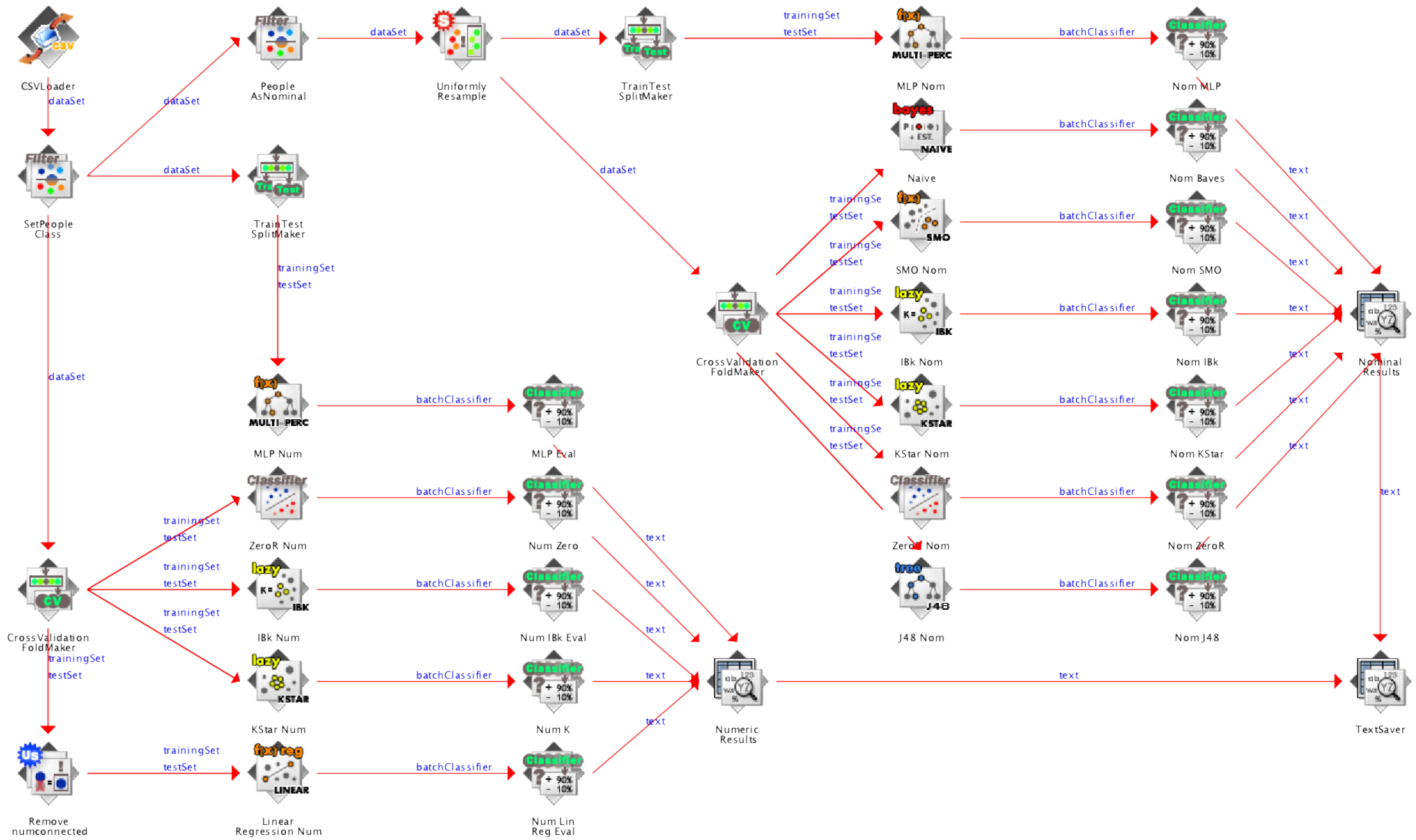


Figure 4.2: Unified Knowledge Flow for all experiments

4.1.5 Classifier Experiment Set 1 Setup

In our first set of experiments, a scene was devised in accordance with Figure 4.3 on page 42 that attempted to sense people from above, as did Thermosense. The prototype was set up on the ceiling, pointing down at a slight angle. For ease of use, the prototype was powered by mains power, and was networked with a laptop for command input and data collection via Ethernet. This set of experiments involved between zero and three people being present in the scene, moving in and out in various ways in accordance with the script in Table 4.2 on the following page.

1. (Remained standing) One person walks in, stands in center, walks out of frame. (sub-experiment 1)

2. (Remained standing) One person walks in, joined by another person, both stand there, one leaves, then another leaves. (sub-experiment 2)

3. (Remained standing) One person walks in, joined by one, joined by another, all stand there, one leaves, then another, then another. (sub-experiment 3)

4. (Remained standing) Two people walk in simultaneously, both stand there, both leave simultaneously. (sub-experiment 4)

5. (Sitting) One person walks in, sits in center, moves to right, walks out of frame. (sub-experiment 5)

6. (Sitting) One person walks in, joined by another person, both sit there, they stand and switch chairs, one leaves, then another leaves. (sub-experiment 6)

7. (Sitting) One person walks in, joined by one, joined by another, they all sit there, one leaves, one shuffles position, then another leaves, then another. (x2) (sub-experiment 7, 8)

8. (Sitting) Two people walk in, both sit there, both leave. (sub-experiment 9)

Table 4.2: Experiment Set 1 Script

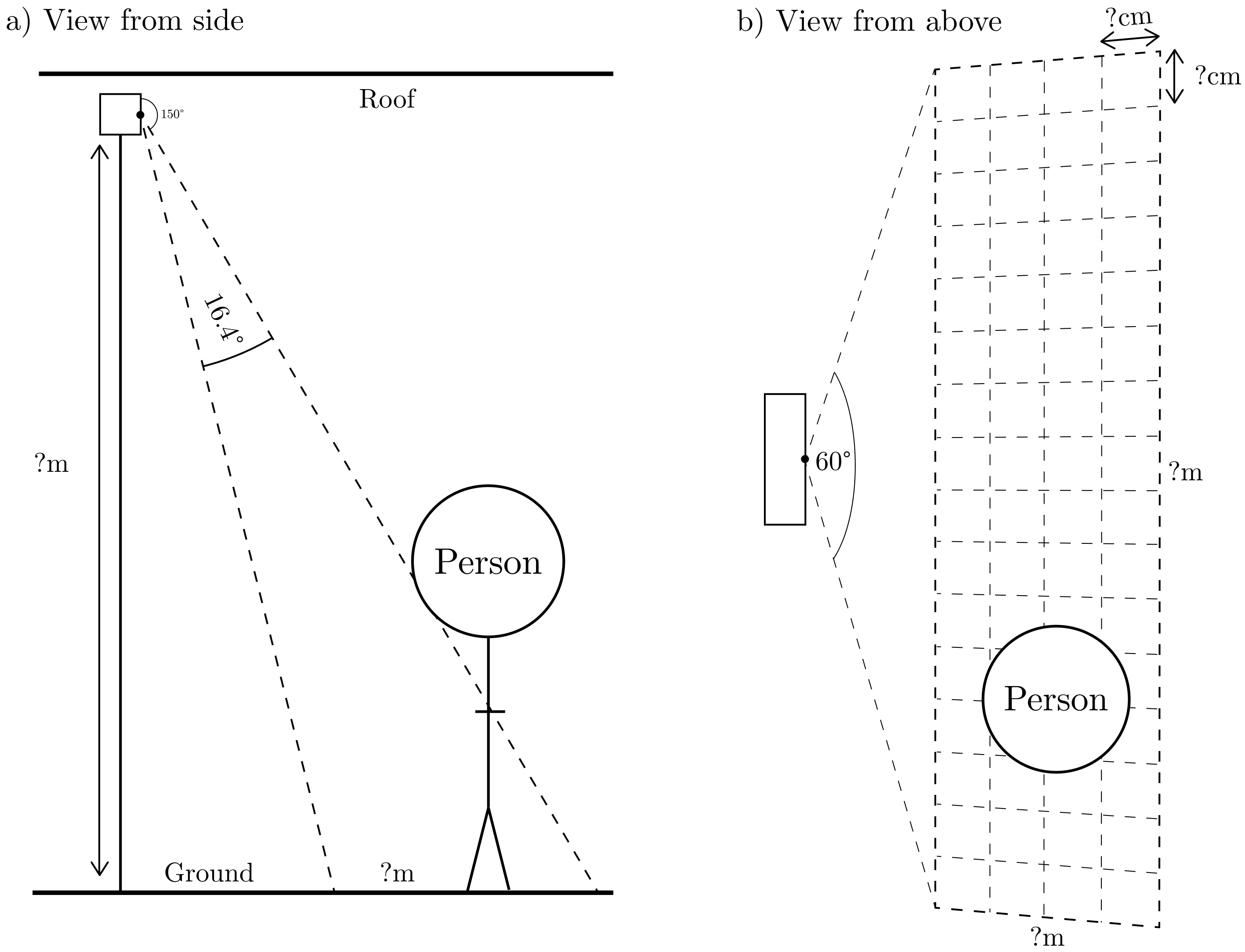


Figure 4.3: Classifier Experiment Set 1 Setup

In these experiments people moved slowly and deliberately, making sure there were large pauses between changes of action. The people involved were of average height, wearing various clothing. The room was cooled to 18 degrees for these experiments.

Each experiment was recorded with a thermal-visual synchronization at 1Hz over approximately 60-120 second intervals. Each experiment had 10-15 frames at the beginning where nothing was within the view of the sensor to allow the thermal background to be calculated. Each frame generated from these experiments was manually tagged with the ground truth value of its occupancy using the script mentioned previously.

The resulting features and ground truth were combined and exported to CSV allowing the Weka machine learning program to analyze them. This data was analyzed with the feature vectors always being considered numeric data and with the ground truth considered both numeric and nominal. All previously mentioned classification algorithms were run against the data set.

4.2 Energy Efficiency Testing

TODO: This section will contain experimental descriptions relating to the power consumption of the prototype.

CHAPTER 5

Results

5.1 Classifier Experiment Set 1

Experimental results from the first set of experiments were overall excellent, results from them can be seen in Table 5.1 on the next page. In the unbalanced results when including zero, an accuracy of 82.9% was observed, and in the balanced results, 78.2%. When excluding zero, the unbalanced results achieved 82.5% in the unbalanced and up to 84.9% in the balanced results.

Between the unbalanced and balanced classes when including zero, the ranking of different algorithms remained approximately the same, and consistently dropped in accuracy, with the exceptions being the SVM technique, which increased in accuracy by about 1% in that instance. The drop in accuracy can be explained mostly by an over-representation of the zero class within the under-balanced data, as well as an underrepresentation of the three class (see Figure 5.1 on page 46). This is conformed by the fact that in the zero-excluded data, there is much less difference in the balanced and unbalanced set. These biases would enable classes to over-predict and under-predict these two classes respectively and achieve an artificially higher accuracy as a result. As discussed in the Methods, we performed re-sampling inside Weka to compensate for this.

For the numeric representation of the number of people, accuracy was consistently poor. From this data, we can see that all three classifiers used performed consistently poorly, with the Root Mean Square Errors being consistently double or more of comparable nominal results, and with correlation coefficients (R^2) indicating poor (or in the case of KNN) very poor correlations.

The two highest accuracy classifiers, C4.5 and K*, achieved quite similar results for both the balanced and unbalanced data while being quite different in implementation.

TODO: Discuss and compare this to ZeroR's RMSE. It's RMSE is quite close to some results, is this bad?

Classifier	**RMSE**		%		R^2	
	Excl. 0	**Incl. 0**	**Excl. 0**	**Incl. 0**	**Excl. 0**	**Incl. 0**
Thermosense Replication						
KNN (Nominal)	-	0.364	-	65.65	-	-
KNN (Numeric)	-	1.1235	-	-	-	0.3766
MLP	0.592	-	-	-	0.687	-
Lin Reg	0.525	-	-	-	0.589	-
Nominal Balanced						
C4.5	0.289	0.290	84.96	77.56	-	-
K*	0.296	0.293	84.27	78.25	-	-
MLP	0.359	0.320	70.86	72.13	-	-
Bayes	0.409	0.368	63.76	61.52	-	-
SVM	0.410	0.386	65.98	56.89	-	-
0-R	0.471	0.433	32.48	24.61	-	-
Numeric						
K*	0.423	0.550	-	-	0.760	0.828
0-R	0.651	0.972	-	-	-0.118	-0.129
Nominal Unbalanced						
C4.5	0.314	0.288	82.39	82.91	-	-
K*	0.304	0.285	82.56	82.61	-	-
MLP	0.362	0.286	77.14	78.69	-	-
Bayes	0.405	0.352	63.59	66.21	-	-
SVM	0.398	0.380	67.18	57.47	-	-
0-R	0.442	0.415	49.74	40.37	-	-

Table 5.1: Classifier Experiment Set 1 Results

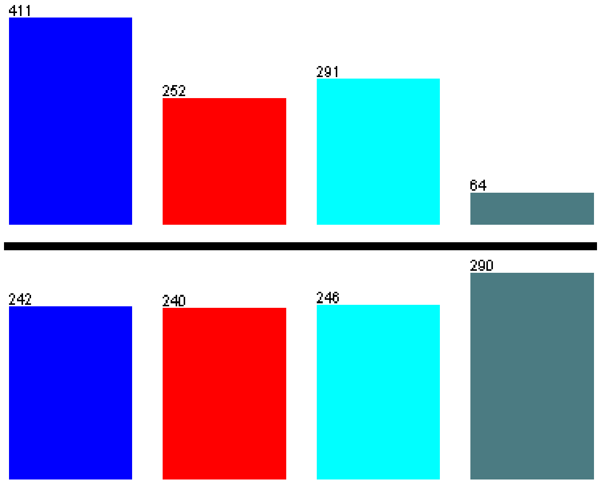


Figure 5.1: Experiment Set 1 Class Distribution Before and After Weka Resampling

5.1.1 Individual sub-experiment results

In addition to the above aggregate classification results, in which each of the nine sub-experiments results are combined and fed into the classifier, each of the sub-experiments has been individually classified with each of the six balanced nominal classifiers above. The results for these classifications can be seen in Table **??** on page ?? and Table 5.2 on the following page.

TODO: We talk about points of interest in the sub-experiment results and see if we can draw any useful conclusions from them.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Avg**
Nominal										
KNN	0.584	0.422	0.375	0.425	0.651	0.352	0.396	0.338	0.244	*0.421*
C4.5	0.342	0.270	0.278	0.414	0.456	0.267	0.318	0.288	0.250	*0.320*
K*	0.305	0.249	0.260	0.412	0.407	0.245	0.299	0.265	0.196	*0.293*
MLP	0.345	0.275	0.272	0.399	0.466	0.232	0.389	0.246	0.242	*0.318*
Bayes	0.391	0.359	0.290	0.435	0.473	0.276	0.381	0.306	0.243	*0.351*
SVM	0.447	0.447	0.379	0.444	0.602	0.378	0.380	0.377	0.335	*0.421*
Numeric										
0-R	0.500	0.471	0.433	0.472	0.500	0.471	0.433	0.433	0.472	*0.465*
KNN	0.726	0.707	1.044	0.934	0.593	0.531	0.899	0.585	0.829	*0.761*
K*	0.256	0.422	0.598	0.806	0.377	0.432	0.586	0.627	0.472	*0.508*
Lin Reg	0.270	0.635	0.623	0.821	0.422	0.508	0.589	0.708	0.570	*0.572*
MLP	0.379	0.460	0.500	0.868	0.399	0.306	0.790	0.490	0.406	*0.511*
0-R	0.409	0.762	1.009	0.986	0.507	0.829	1.014	1.043	1.012	*0.841*
Avg	*0.413*	*0.457*	*0.505*	*0.618*	*0.488*	*0.402*	*0.539*	*0.475*	*0.439*	

Table 5.2: Classifier Experiment Set 1 Individual Sub-experiment RMSEs

5.2 Thermosense Comparison

To aid in comparison with the Thermosense algorithm, in our experiment sets we chose three techniques similar to those used in Thermosense; an Artificial Neural Network (Multilayer Perceptron in Weka), k-nearest Neighbors (IBk in Weka) and Linear Regression. Our results for our approximations to their algorithms can be seen in the Thermosense Replication section of Table 5.1 on page 45. We also use an experimental setup with 0–3 or 1–3 people, depending on the classifier, just as the Thermosense experiments do. For KNN, it was ambiguous if Thermosense used numeric or a nominal class attributes in their experiments, so we have performed both options above. The specifics of this is discussed in the Methods chapter.

For those experiments where we tried to specifically replicate Thermosense' results, we found consistently poor results that did not come close to meeting the R^2 or RMSE values that Thermosense claimed. Their R^2 values were in the 0.8 range, while ours range from around 0.3–0.7. This can be put down to a multitude of factors, with one likely explanation being that differences with the Melexis MLX90620 (*Melexis*)'s narrower field of view being difficult for Thermosense' specific algorithmic selections to deal with.

0.141 S -0.051 0.201 F-statistic 3220 R2 0.858 2.44 10187 1.14 1030 9.25 1011

Our Linear Regression (Equation (5.1)) underperforms particularly when compared to Thermosense' (Equation (5.2)), as it fails to find an adequate way to weight the two variables, instead opting to basically exclude them from consideration by picking very small weights and adding a large constant factor. We get a correlation of $R^2 = 0.589$ vs Thermosense' $R^2 = 0.858$.

$$n = 0.0456a + -0.024s + 1.1772 \tag{5.1}$$

$$n = 0.141a + -0.051s + 0.201 \tag{5.2}$$

However, for those algorithms we chose ourselves to test, we found results that were comparable or even better than Thermosense. Our nominally classed C4.5 decision tree achieved an balanced RMSE excluding zero of 0.289, compared to Thermosense' best result of 0.346. In the numeric classes, the K* implementation achieved correlations in the same ballpark of 0.760 (excl. 0) and 0.828 (incl. 0), compared to Thermosense' best 0.906 for their ANN.

However, none of the techniques used by Thermosense proved to be the best out of those algorithms tried. The Neural Network and k-nearest Neighbors techniques represent the middle-of-the-road of our results, with both being bested by the C4.5 and K* algorithms, which produced RMSEs of 0.289 and 0.298 respectively. Both these results are significantly better than Thermosense's best RMSE of 0.346 for k-nearest Neighbors, with our C4.5 algorithm representing a 28% improvement over that technique.

CHAPTER 6

Discussion and Conclusion

The smart-home economy continues to grow, with automation being one of the main areas driving growth. The ability to detect people present within a space is an important smart-automation feature, with the implications for climate control energy efficiency alone being highly significant.

This project has attempted to create an occupancy detection system for such a smart home environment that meets four criteria; Low Cost, Non-Invasive, Energy Efficient and Reliable. Building such a system to commercial standards is outside of the scope of this project, however a prototype that attempts to prove the concepts involved was built and tested against these criteria. This prototype was based upon the ceiling-mounted thermal imaging approach of Thermosense [7], which after extensive analysis proved to be the best option given our criteria.

This prototype both validates the methods and results of the Thermosense paper, discovers key caveats surrounding the Thermosense approach, and also creates a software and hardware base on which future research into the area of occupancy in thermal imaging can be explored.

6.1 Criteria

6.1.1 Low Cost

As being low-cost is one of the project's goals, we have summarized the cost of each of the components of the prototype in Table 6.1 on the following page. We believe that for a prototype, this cost is sufficiently low. In an ideal system, there would only be one Raspberry Pi in the system, and it would not require a camera, lowering the cost to around $\$40 + n * \$115$ where n is the number of sensors. Similarly, as technology improves (as discussed in later chapters), sensor technology will continue to become cheaper, causing the most expensive component, the Melexis MLX90620 (*Melexis*), to lower in cost.

Part	Cost (USD)
MLX90620	$80
Raspberry Pi B+	$40
Raspberry Pi Camera	$30
Arduino Uno R3	$25
Passive Infrared Sensor	$10
TOTAL	**$185**

Table 6.1: Our project component costs

Part	Cost (USD)
Tmote Sky	$100
Grid-EYE	$30
Passive Infrared Sensor	$10
TOTAL	**$140**

Table 6.2: Thermosense component costs

When we compare this to the estimated cost of the Thermosense system (Table 6.2), we believe that it achieves a suitably comparible cost for a prototype. When removing the aspects of the prototype that would be unnecessary in the final version, the difference is less than $15.

6.1.2 Non-Invasive

As discussed in the Literature Review, low-resolution thermal sensing provides the best trade-off between accuracy and invasiveness. Due to sensing in the infrared spectrum, it becomes significantly harder to surveil people in a malicious way, as many identifying features of people are not visible in the IR spectrum. This is compounded by the low resolution, which similarly assists in reducing the invasiveness of the sensor.

6.1.3 Energy Efficient

6.1.4 Reliable

As discussed in the Results section, the prototype developed achieves excellent reliability, citing accuracies in the 80% range.

6.2 Future Directions

This project merely touched upon the area of thermal sensing and occupancy detection, and has laid the foundation for many more projects that build upon this original project. Some areas of future research are discussed here;

6.2.1 Sub-pixel localization

Due to the overlapping bell-curve characteristics of the Melexis MLX90620 (*Melexis*)'s pixels, it may be possible to perform sub-pixel localization on objects within images.

6.2.2 Improving Robustness

One of the main areas of the project that was not explored due to time was the introducting of a wireless mesh networking architecture to the project. Future prototypes would consist of an many-to-one relationship between the Sensing/Preprocessing tier and the Analysis tier. Exploring the best way to mesh network these components while maintaining all the pre-existing criteria of the project would be challenging. In Appendix Chapter B on page 61 we provide our thoughts on the potential structure this could take.

Similarly, the current prototype uses a breadboarded structure that increases the size of the prototype significantly, as well as reduces the reliability of the prototype in the long-term. Converting the *Melexis* and PIR into a printed circuit board that fits onto the Arduino as a shield would both reduce the size of the prototype, as well as improving reliability for the future.

6.2.3 Field-of-view modifications

Several different techniques could be used to improve upon the field-of-view limitations of the *Melexis*, and exploring them and their cost/complexity implications would be useful. The first of these is applying a lens to the sensor, effectively expanding the field-of-view, but at the cost of distorting the image. Compensating for this distortion while maintaining accuracy presents an intriguing problem.

In another direction, using a motor with the *Melexis* to "sweep" the room, and thereby constructing a larger image of the space could also resolve the field-of-view issues. However, this approach also presents problems in stitching the images together in a sensible way, the distortion caused by rotating the sensor, as well as handing cases in which a fast-moving object is represented multiple times in the stitched image.

6.2.4 New Sensors

During this project, an updated version of our sensor, the MLX90621, was released. This version doubles the field-of-view in both the horizontal and vertical directions, addressing many of the problems encountered with the size of detection area in low-ceiling rooms. This version offers nearly complete backwards compatibility with the older version. Updating the project code-base to support it and re-running the experiments with the increased field-of-view to determine how much of an improvement it is would be interesting.

In addition to this, significantly higher resolution sensors are beginning to come to the market. The FLiR Lepton [24], which sells in a dev kit for $350, offers an 80×60 pixel sensor with a comparable field-of-view to the Grid-EYE. Exploring the increases in accuracy achievable though such significant increases in resolution would have significant contrion.

APPENDIX A

Original Honours Proposal

Title:	Developing a robust system for occupancy detection in the household
Author:	Ash Tyndall
Supervisor:	Professor Rachel Cardell-Oliver
Degree:	BCompSci (24 point project)
Date:	October 8, 2014

A.1 Background

The proportion of elderly and mobility-impaired people is predicted to grow dramatically over the next century, leaving a large proportion of the population unable to care for themselves, and consequently less people able care for these groups. [5] With this issue looming, investments are being made into a variety of technologies that can provide the support these groups need to live independent of human assistance.

With recent advancements in low cost embedded computing, such as the Arduino [1] and Raspberry Pi, [14] the ability to provide a set of interconnected sensors, actuators and interfaces to enable a low-cost 'smart home for the disabled' is becoming increasingly achievable.

Sensing techniques to determine occupancy, the detection of the presence and number of people in an area, are of particular use to the elderly and disabled. Detection can be used to inform various devices that change state depending on the user's location, including the better regulation energy hungry devices to help reduce financial burden. Household climate control, which in some regions of Australia accounts for up to 40% of energy usage [2] is one particular area in which occupancy detection can reduce costs, as efficiency can be increased dramatically with annual energy savings of up to 25% found in some cases. [7]

Significant research has been performed into the occupancy field, with a focus on improving the energy efficiency of both office buildings and households. This is achieved through a variety of sensing means, including thermal arrays, [4] ultrasonic sensors, [10] smart phone tracking, [11][3] electricity consumption, [12] network traffic analysis, [15] sound, [9] CO2, [9] passive infrared, [9] video cameras, [6] and various fusions of the above. [16][15]

A.2 Aim

While many of the above solutions achieve excellent accuracies, in many cases they suffer from problems of installation logistics, difficult assembly, assumptions on user's technology ownership and component cost. In a smart home for the disabled, accuracy is important, but accessibility is paramount.

The goal of this research project is to devise an occupancy detection system that forms part of a larger 'smart home for the disabled' that meets the following accessibility criteria;

- *Low Cost*: The set of components required should aim to minimise cost, as these devices are intended to be deployed in situations where the serviced user may be financially restricted.
- *Non-Invasive*: The sensors used in the system should gather as little information as necessary to achieve the detection goal; there are privacy concerns with the use of high-definition sensors.
- *Energy Efficient*: The system may be placed in a location where there is no access to mains power (i.e. roof), and the retrofitting of appropriate power can be difficult; the ability to survive for long periods on only battery power is advantageous.
- *Reliable*: The system should be able to operate without user intervention or frequent maintenance, and should be able to perform its occupancy detection goal with a high degree of accuracy.

Success in this project would involve both

1. Devising a bill of materials that can be purchased off-the-shelf, assembled without difficulty, on which a software platform can be installed that performs analysis of the sensor data and provides a simple answer to the occupancy question, and
2. Using those materials and softwares to create a final demonstration prototype whose success can be tested in controlled and real-world conditions.

This system would be extensible, based on open standards such as REST or CoAP, [8][13] and could easily fit into a larger 'smart home for the disabled' or internet-of-things system.

A.3 Method

Achieving these aims involves performing research and development in several discrete phases.

A.3.1 Hardware

A list of possible sensor candidates will be developed, and these candidates will be ranked according to their adherence to the four accessibility criteria outlined above. Primarily the sensor ranking will consider the cost, invasiveness and reliability of detection, as the sensors themselves do not form a large part of the power requirement.

Similarly, a list of possible embedded boards to act as the sensor's host and data analysis platform will be created. Primarily, they will be ranked on cost, energy efficiency and reliability of programming/system stability.

Low-powered wireless protocols will also be investigated, to determine which is most suitable for the device; providing enough range at low power consumption to allow easy and reliable communication with the hardware.

Once promising candidates have been identified, components will be purchased and analysed to determine how well they can integrate.

A.3.2 Classification

Depending on the final sensor choice, relevant experiments will be performed to determine the classification algorithm with the best occupancy determination accuracy. This will involve the deployment of a prototype to perform data gathering, as well as another device/person to assess ground truth.

A.3.3 Robustness / API

Once the classification algorithm and hardware are finalised, an easy to use API will be developed to allow the data the device collects to be integrated into a broader system.

The finalised product will be architected into a easy-to-install software solution that will allow someone without domain knowledge to use the software and corresponding hardware in their own environment.

A.4 Timeline

Date	Task
Fri 15 August	*Project proposal and project summary due to Coordinator*
August	Hardware shortlisting / testing
25–29 August	*Project proposal talk presented to research group*
September	Literature review
Fri 19 September	*Draft literature review due to supervisor(s)*
October - November	Core Hardware / Software development
Fri 24 October	*Literature Review and Revised Project Proposal due to Coordinator*
November - February	*End of year break*
February	Write dissertation
Thu 16 April	*Draft dissertation due to supervisor*
April - May	Improve robustness and API
Thu 30 April	*Draft dissertation available for collection from supervisor*
Fri 8 May	*Seminar title and abstract due to Coordinator*
Mon 25 May	*Final dissertation due to Coordinator*
25–29 May	*Seminar Presented to Seminar Marking Panel*
Thu 28 May	*Poster Due*
Mon 22 June	*Corrected Dissertation Due to Coordinator*

A.5 Software and Hardware Requirements

A large part of this research project is determining the specific hardware and software that best fit the accessibility criteria. Because of this, an exhaustive list of software and hardware requirements are not given in this proposal.

A budget of up to $300 has been allocated by my supervisor for project purchases. Some technologies with promise that will be investigated include;

Raspberry Pi Model B+ Small form-factor Linux computer
Available from `http://arduino.cc/en/Guide/Introduction`; $38

Arduino Uno Small form-factor microcontroller
Available from `http://arduino.cc/en/Main/arduinoBoardUno`; $36

Panasonic Grid-EYE Infrared Array Sensor
Available from `http://www3.panasonic.biz/ac/e/control/sensor/infrared/grid-eye/index.jsp`; approx. $33

Passive Infrared Sensor
Available from various places; $10–$20

A.6 Proposal References

[1] Ardunio. `http://arduino.cc/en/Guide/Introduction`. Accessed: 2014-08-09.

[2] Australian Bureau of Statistics. Household water and energy use, Victoria: Heating and cooling. Tech. Rep. 4602.2, October 2011. Retrieved October 6, 2014 from `http://www.abs.gov.au/ausstats/abs@.nsf/0/85424ADCCF6E5AE9CA257A670013AF89`.

[3] Balaji, B., Xu, J., Nwokafor, A., Gupta, R., and Agarwal, Y. Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (2013), ACM, p. 17.

[4] Beltran, A., Erickson, V. L., and Cerpa, A. E. ThermoSense: Occupancy thermal based sensing for HVAC control. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–8.

[5] Chan, M., Campo, E., Estève, D., and Fourniols, J.-Y. Smart homes - current features and future perspectives. *Maturitas 64*, 2 (2009), 90–97.

[6] Erickson, V. L., Achleitner, S., and Cerpa, A. E. POEM: Power-efficient occupancy-based energy management system. In *Proceedings of the 12th international conference on Information processing in sensor networks* (2013), ACM, pp. 203–216.

[7] Erickson, V. L., Beltran, A., Winkler, D. A., Esfahani, N. P., Lusby, J. R., and Cerpa, A. E. Demo abstract: ThermoSense: thermal array sensor networks in building management. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (2013), ACM, p. 87.

[8] Guinard, D., Ion, I., and Mayer, S. In search of an internet of things service architecture: REST or WS-*? a developers perspective. In *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 2012, pp. 326–337.

[9] Hailemariam, E., Goldstein, R., Attar, R., and Khan, A. Real-time occupancy detection using decision trees with multiple sensor types. In *Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design* (2011), Society for Computer Simulation International, pp. 141–148.

[10] Hnat, T. W., Griffiths, E., Dawson, R., and Whitehouse, K. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* (2012), ACM, pp. 309–322.

[11] Kleiminger, W., Beckel, C., Dey, A., and Santini, S. Poster abstract: Using unlabeled Wi-Fi scan data to discover occupancy patterns of private households. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (2013), ACM, p. 47.

[12] Kleiminger, W., Beckel, C., Staake, T., and Santini, S. Occupancy detection from electricity consumption data. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–8.

[13] Kovatsch, M. CoAP for the web of things: from tiny resource-constrained devices to the web browser. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (2013), ACM, pp. 1495–1504.

[14] Raspberry pi. `http://www.raspberrypi.org/`. Accessed: 2014-08-09.

[15] Ting, K., Yu, R., and Srivastava, M. Poster Abstract: Occupancy inferencing from non-intrusive data sources. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–2.

[16] Yang, Z., Li, N., Becerik-Gerber, B., and Orosz, M. A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations. In *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design* (2012), Society for Computer Simulation International, p. 2.

APPENDIX B

Ideal System Architecture

Beyond specific sensor design and occupancy detection algorithms, a core goal of this project is to create a system that is designed to operate as a useful Thing in a real-world Internet of Things (IoT) environment, as the key advantage of Things is the “disruptive level of innovation”[4] brought about by their ability to be combined in ways unforeseen (yet still enabled) by their creators. This architecture involves careful consideration of the embedded hardware that will drive the system, as well as the communications protocols utilised between the sensor and devices interested in the sensor's information.

B.1 Protocols

In an ideal smart-home environment, the sensor systems used will communicate with each other wirelessly. As the complete sensor system has low power requirements to enable battery operation, it is important to prioritise those protocols and architectures that minimise power usage while still enabling the necessary wireless communication. The system will also ideally exist in a system with other identical sensors (one for each room in a residence), thus it is important to prioritise those protocols which allow multiple identical sensor systems to coexist on the same network without conflict, and to be uniquely addressable and iden-

REST	
Application	CoAP
Transport	UDP
IP / Routing	IETF RPL
Adaptation	IETF 6LoWPAN
Medium Access	IEEE 802.15.4e
Physical	IEEE 802.15.4-2006

Table B.1: Proposed protocol stack

tifiable. In recent years, many developments have been made in the Internet of Things (IoT) arena, with standards emerging specifically designed for low-power embedded devices to communicate between themselves and bigger systems that address these and other unique needs, across the entire protocol stack.

Palattella et al. [21] propose a protocol stack that aligns with the above requirements, with the key advantage being a wholly standardized implementation of the stack exists. This implementation is based on TCP/IP, uses the latest IEEE and IETF IoT standards, and is free from proprietary protocol restrictions (unlike ZigBee 1.0 devices, for instance). Table B.1 on the previous page shows the full stack proposed. The key components of this proposal are the introduction of CoAP at the application layer, RPL at the IP / Routing layer and 6LoWPAN at the Adaptation layer.

Above the application layer, Guinard et al. [11] propose the use of Representational state transfer (REST) over Web Services Descriptive Language / Simple Object Access Protocol (WS-*) as a method of exchanging information between sensor systems. Their data suggests that REST is easier to use than WS-*, and the key advantage of a WS-* based approach is its ability to represent much more complex data and abstractions, which are unnecessary in this project's situation.

Constrained Application Protocol (CoAP) [18] is an application layer protocol designed to replace HTTP as a way of transmitting RESTful information between clients. The chief advantage of CoAP over HTTP is it compresses the broadstrokes of the HTTP feature set into a binary language that is much more suitable for transmission over low-bandwidth and low-power links, such as those discussed here.

IPv6 Routing Protocol for Low-Power and Lossy Networks (RPL) [27] is a routing protocol designed for low power environments, allowing low power nodes to create and maintain a mesh network between themselves, allowing, among other things, the routing of packets to a "root" node and back again. RPL is particularly suited to the routing situation of our proposed architecture, as individual sensors do not need to communicate with one another, but rather report back to a larger node (further discussed in Subsection B.2 on the following page).

IPv6 over Low power Wireless Personal Area Networks (6LoWPAN) [23] is a compression and formatting specification to allow IPv6 packets to be sent over an 802.15.4 based network. Optimisations are found in the reduction of the size of 6LoWPAN packets, IPv6 addresses as well as redesigning core Internet Protocol algorithms so that they can run with low power consumption on participating devices.

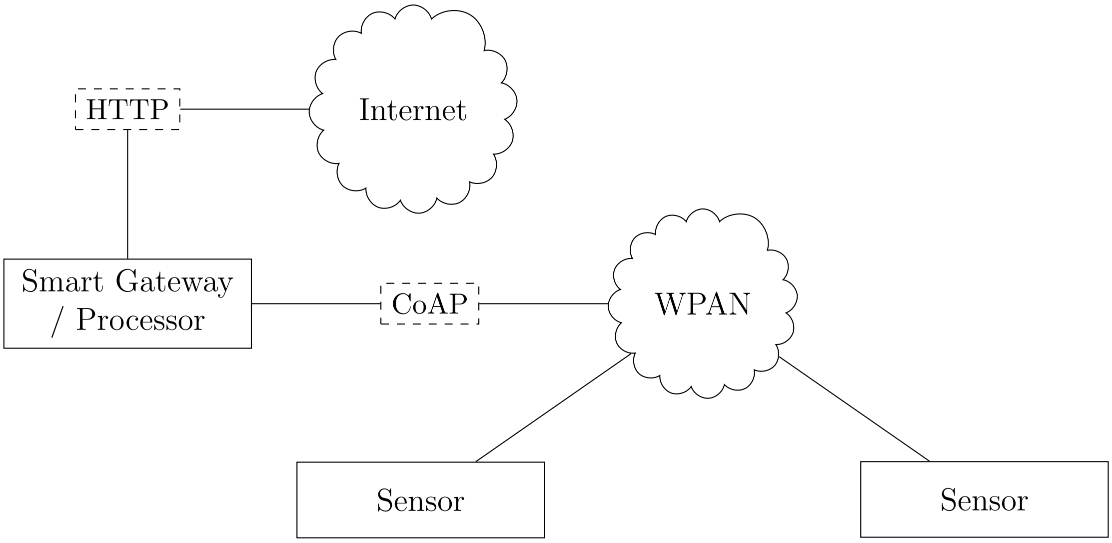


Figure B.1: Proposed system architecture

B.2 Devices

In addition to the protocol stack used, how these nodes relate to each other is also an important consideration. Part of what will inform these decisions are the requisite processing power and internet connectivity required to successfully execute all elements of the sensing system. Kovatsch [18] provides a constructive classification system to consider this, by describing three classes of resource constrained devices that would benefit from Constrained Application Protocol (CoAP), and each can provide different levels of security for an IP stack;

- *Class 0*: "not capable of running an RFC-compliant IP stack in a secure manner. They require application-level gateways to connect to the Internet."
- *Class 1*: Able to connect to the internet with some "integrated security mechanisms". Are unable to employ full HTTP with TLS.
- *Class 2*: Normal Internet nodes, able to use the full HTTP stack with TLS.

The devices that we propose the sensors will connect to are the likes of the Arduino, which can be classified as class 0 or possibly class 1 devices. Due to their insecurity and difficulty running a fully fledged IP stack, Guinard et al. [12] propose the use of a "Smart Gateway" system to bridge the wider internet and these sensor systems. This gateway would be able to communicate with the sensor systems over CoAP and 802.15.4 , as well as receive API requests via HTTP from a traditional TCP/IP network to forward on to these sensors.

The Thermosense paper [7] proposes several different algorithms to process the raw sensing data into the occupancy estimates (further discussed in Section 2.4 on page 10), all of which are fairly computationally expensive. Because of this, it would be non-trivial to implement these algorithms on the embedded sensing devices themselves. This problem is already resolved in our proposed system, as the aforementioned "Smart Gateway" can easily also take on the task of processing the raw sensor data into estimates which it can relay to interested parties over its HTTP-based API. A visualisation of this proposed system is shown in Figure B.1 on the preceding page.

APPENDIX C

Code Listings

C.1 ThingLib

C.1.1 cam.py

```
from __future__ import division
from __future__ import print_function

import serial
import copy
import Queue as queue
import time
from collections import deque
import threading
import pygame
import colorsys
import datetime
from PIL import Image, ImageDraw, ImageFont
import subprocess
import tempfile
import os
```

```python
import os.path
import fractions
import pxdisplay
import multiprocessing
import numpy as np
import io



class BaseManager(object):
  driver = None
  build = None
  irhz = None

  tty = None
  baud = None

  hflip = True
  vflip = True

  _temps = None
  _serial_obj = None
  _queues = []

  def __init__(self, tty, hz=8, baud=115200, init=True):
    self.tty = tty
    self.baud = baud
    self.irhz = hz

    if init:
      self._serial_obj = serial.Serial(port=self.tty, baudrate=self.baud, rtscts=True, dsrdtr=True)

  def __del__(self):
    self.close()
```

```python
def _reset_and_conf(self, timers=True):
  self._serial_obj.write('r\n') # Reset the sensor
  self._serial_obj.flush()

  time.sleep(2)

  if timers:
    self._serial_obj.write('t\n') # Turn on timers
  else:
    self._serial_obj.write('o\n') # Turn on timers

  self._serial_obj.flush()

def _decode_packet(self, packet, splitchar="\t"):
  decoded_packet = {}
  ir = []

  for line in packet:
    parted = line.partition(" ")
    cmd = parted[0]
    val = parted[2]

    try:
      if cmd == "START":
        decoded_packet['start_millis'] = long(val)
      elif cmd == "STOP":
        decoded_packet['stop_millis'] = long(val)
      elif cmd == "MOVEMENT":
        if val == "0":
          decoded_packet['movement'] = False
        elif val == "1":
          decoded_packet['movement'] = True
      else:
```

```python
        ir.append(tuple(float(x) for x in line.split(splitchar)))
      except ValueError:
        print(packet)
        print("WARNING: Could not decode corrupted packet")
        return {}

  if self.hflip:
    ir = map(tuple, np.fliplr(ir))

  if self.vflip:
    ir = map(tuple, np.flipud(ir))

  decoded_packet['ir'] = tuple(ir)

  return decoded_packet

def _decode_info(self, packet):
  decoded_packet = {}
  ir = []

  for line in packet:
    parted = line.partition(" ")
    cmd = parted[0]
    val = parted[2]

    if cmd == "INFO":
      pass
    elif cmd == "DRIVER":
      decoded_packet['driver'] = val
    elif cmd == "BUILD":
      decoded_packet['build'] = val
    elif cmd == "IRHZ":
      decoded_packet['irhz'] = int(val) if int(val) != 0 else 0.5
```

```python
    return decoded_packet

  def _update_info(self):
    ser = self._serial_obj

    ser.write('i')
    ser.flush()
    imsg = []

    line = ser.readline().decode("ascii", "ignore").strip()

    # Capture a whole packet
    while not line == "INFO START":
      line = ser.readline().decode("ascii", "ignore").strip()

    while not line == "INFO STOP":
      imsg.append(line)
      line = ser.readline().decode("ascii", "ignore").strip()

    imsg.append(line)

    packet = self._decode_info(imsg)

    self.driver = packet['driver']
    self.build = packet['build']

    if packet['irhz'] != self.irhz:
      ser.write('f{}'.format(self.irhz))
      self._update_info()

  def _wait_read_packet(self):
    ser = self._serial_obj
    line = ser.readline().decode("ascii", "ignore").strip()
    msg = []
```

```
  # Capture a whole packet
  while not line.startswith("START"):
    line = ser.readline().decode("ascii", "ignore").strip()

  while not line.startswith("STOP"):
    msg.append(line)
    line = ser.readline().decode("ascii", "ignore").strip()

  msg.append(line)

  return msg

def close(self):
  return

def get_temps(self):
  if self._temps is None:
    return False
  else:
    return copy.deepcopy(self._temps)

def subscribe(self):
  q = queue.Queue()
  self._queues.append(q)
  return q

def subscribe_multiprocess(self):
  q = multiprocessing.Queue()
  self._queues.append(q)
  return q

def subscribe_lifo(self):
  q = queue.LifoQueue()
```

```python
    self._queues.append(q)
    return q



class Manager(BaseManager):
  _serial_thread = None
  _serial_stop = False
  _serial_ready = False

  _decode_thread = None

  _read_decode_queue = None

  def __init__(self, tty, hz=8, baud=115200):
    super(self.__class__, self).__init__(tty, hz, baud)

    self._serial_thread = threading.Thread(group=None, target=self._read_thread_run)
    self._serial_thread.daemon = True

    self._decode_thread = threading.Thread(group=None, target=self._decode_thread_run)
    self._decode_thread.daemon = True

    self._reset_and_conf(timers=True)

    self._read_decode_queue = queue.Queue()

    self._decode_thread.start()
    self._serial_thread.start()

    while not self._serial_ready: # Wait until we've populated data before continuing
      pass

  def close(self):
```

```
    self._serial_stop = True

    if self._serial_thread is not None:
      while self._serial_thread.is_alive(): # Wait for thread to terminate
        pass

  def _read_thread_run(self):
    ser = self._serial_obj
    q = self._read_decode_queue
    self._update_info()

    while True:
      msg = self._wait_read_packet()

      q.put(msg)
      self._serial_ready = True

      if self._serial_stop:
        ser.close()
        return

  def _decode_thread_run(self):
    dq = self._read_decode_queue
    while True:
      msg = dq.get(block=True)

      dpct = self._decode_packet(msg)

      if 'ir' in dpct:
        self._temps = dpct

        for q in self._queues:
          q.put(self.get_temps())
```

```
      if self._serial_stop:
        return


class OnDemandManager(BaseManager):
  def __init__(self, tty, hz=8, baud=115200):
    super(self.__class__, self).__init__(tty, hz, baud)

    self._reset_and_conf(timers=False)

    self._update_info()

  def close(self):
    self._serial_obj.close()

  def capture(self):
    self._serial_obj.write('p') # Capture frame manually
    self._serial_obj.flush()

    msg = self._wait_read_packet()
    dpct = self._decode_packet(msg)

    if 'ir' in dpct:
      self._temps = dpct

      for q in self._queues:
        q.put(self.get_temps())

    return dpct



class ManagerPlaybackEmulator(BaseManager):
  _playback_data = None
```

```
_pb_thread = None
_pb_stop = False
_pb_len = 0

_i = 0

def __init__(self, playback_data=None):
  if playback_data is not None:
    self.irhz, self._playback_data = playback_data
    self._pb_len = len(self._playback_data)

  self.driver = "Playback"
  self.build = "1"

def set_playback_data(self, playback_data):
  self.stop()
  self.irhz, self._playback_data = playback_data
  self._pb_len = len(self._playback_data)

def close(self):
  return

def start(self):
  if self._pb_thread is None:
    self._pb_stop = False
    self._pb_thread = threading.Thread(group=None, target=self._pb_thread_run)
    self._pb_thread.daemon = True
    self._pb_thread.start()

def pause(self):
  self._pb_stop = True

  while self._pb_thread is not None and self._pb_thread.is_alive():
```

```
      pass

    self._pb_thread = None

  def stop(self):
    self._pb_stop = True

    while self._pb_thread is not None and self._pb_thread.is_alive():
      pass

    self._pb_thread = None
    self._i = 0

  def get_temps(self):
    return self._playback_data[self._i]

  def _pb_thread_run(self):
    while True:
      if self._pb_stop:
        return

      for q in self._queues:
        q.put(self._playback_data[self._i])

      time.sleep(1.0/float(self.irhz))

      self._i += 1

      if self._i >= self._pb_len:
        return



class Visualizer(object):
```

```
_display_thread = None
_display_stop = False
_tmin = None
_tmax = None
_limit = None
_dwidth = None

_tcam = None
_ffmpeg_loc = None

_camera = None

def __init__(self, tcam=None, camera=None, ffmpeg_loc="ffmpeg"):
  self._tcam = tcam
  self._ffmpeg_loc = ffmpeg_loc
  self._camera = camera

def display(self, block=False, limit=0, width=100, tmin=15, tmax=45):
  q = self._tcam.subscribe_multiprocess()
  _, proc = pxdisplay.create(q, limit=limit, width=width, tmin=tmin, tmax=tmax)

  if block:
    proc.join()

def playback(self, filen, tmin=15, tmax=45):
  hz, playdata = self.file_to_capture(filen)

  print(hz)

  q, thread = pxdisplay.create(
    limit=hz,
    tmin=tmin,
    tmax=tmax,
    caption="Playing back '{}'".format(filen)
```

```
    )

    start = datetime.datetime.now()
    offset = playdata[0]['start_millis']

    for n, frame in enumerate(playdata):
      frame['text'] = 'T+%.3f' % ((frame['start_millis'] - offset)/ 1000.0)
      q.put(frame)

  def display_close(self):
    if self._display_thread is None:
      return

    self._display_stop = True
    self._display_thread = None

  def close(self):
    self.display_close()

  def capture_to_file(self, capture, hz, filen):
    with open(filen + '_thermal.hcap', 'w') as f:
      f.write(str(hz) + "\n")

      for frame in capture:
        t = frame['start_millis']
        motion = frame['movement']
        arr = frame['ir']
        f.write(str(t) + "\n")
        f.write(str(motion) + "\n")
        for l in arr:
          f.write('\t'.join([str(x) for x in l]) + "\n")
        f.write("\n")

  def capture_to_img_sequence(self, capture, directory, tmin=15, tmax=45, text=True):
```

```
  hz, frames = capture
  pxwidth = 120
  print(directory)

  for i, frame in enumerate(frames):
    im = Image.new("RGB", (1920, 480))
    draw = ImageDraw.Draw(im)
    font = ImageFont.truetype("arial.ttf", 35)

    for k, row in enumerate(frame['ir']):
      for j, px in enumerate(row):
        rgb = pxdisplay.temp_to_rgb(px, tmin, tmax)

        x = k*pxwidth
        y = j*pxwidth

        coords = (y, x, y+pxwidth+1, x+pxwidth+1)

        draw.rectangle(coords, fill=rgb)

        if text:
          draw.text([y+20,x+(pxwidth/2-20)], str(px), fill=(255,255,255), font=font)

    im.save(os.path.join(directory, '{:09d}.png'.format(i)))

def capture_to_movie(self, capture, filename, width=1920, height=480, tmin=15, tmax=45):
  hz, frames = capture
  tdir = tempfile.mkdtemp()

  self.capture_to_img_sequence(capture, tdir, tmin=tmin, tmax=tmax)

  args = [self._ffmpeg_loc,
    "-y",
    "-r", str(fractions.Fraction(hz)),
```

```
      "-i", os.path.join(tdir, "%09d.png"),
      "-s", "{}x{}".format(width, height),
      "-sws_flags", "neighbor",
      "-sws_dither", "none",
      '-vcodec', 'qtrle', '-pix_fmt', 'rgb24',
      filename + '_thermal.mov'
      ]

    subprocess.call(args)

  def file_to_capture(self, filen):
    capture = []
    hz = None
    with open(filen + '_thermal.hcap', 'r') as f:
      frame = {'ir':[]}

      for i, line in enumerate(f):
        if i == 0:
          hz = float(line)
          continue

        j = (i-1) % 7
        if j == 0:
          frame['start_millis'] = int(line)
        elif j == 1:
          frame['movement'] = bool(line)
        elif 1 < j < 6:
          frame['ir'].append(tuple([float(x) for x in line.split("\t")]))
        elif j == 6:
          capture.append(frame)
          frame = {'ir':[]}

    return (hz, capture)
```

```
def capture(self, seconds, name=None, hcap=False, video=False):
  buff = []
  q = self._tcam.subscribe()
  hz = self._tcam.irhz
  tdir = tempfile.mkdtemp()

  camera = None
  visfile = name + '_visual.h264' #os.path.join(tdir, name + '_visual.h264')

  if video and self._camera is not None:
    self._camera.resolution = (1920, 1080)
    self._camera.framerate = hz
    self._camera.start_recording(visfile)

  start = time.time()
  elapsed = 0

  while elapsed <= seconds:
    elapsed = time.time() - start
    buff.append( q.get() )

  if video and self._camera is not None:
    self._camera.stop_recording()

    #args = [self._ffmpeg_loc,
    #   "-y",
    #   "-r", str(fractions.Fraction(hz)),
    #   "-i", visfile,
    #   "-vcodec", "copy",
    #   name + '_visual.mp4'
    #   ]

    #subprocess.call(args)

```

```python
    #os.remove(visfile)


  if hcap:
    self.capture_to_file(buff, hz, name)

  return (hz, buff)

def capture_synced(self, seconds, name, hz=2):
  cap_method = getattr(self._tcam, "capture", None)
  if not callable(cap_method):
    raise "Provided tcam class must support the capture method"

  if self._camera is None:
    raise "No picamera object provided, cannot proceed"

  camera = self._camera
  camera.resolution = (1920, 1080)

  # TODO: Currently produces black images. Need to fix.
  # Wait for analog gain to settle on a higher value than 1
  #while camera.analog_gain <= 1 or camera.digital_gain <= 1:
  #    time.sleep(1)

  # Now fix the values
  #camera.shutter_speed = camera.exposure_speed
  #camera.exposure_mode = 'off'
  #g = camera.awb_gains
  #camera.awb_mode = 'off'
  #camera.awb_gains = g

  import datetime, threading, time

  dir_name = name
```

```
frames = seconds * hz

buff = []
imgbuff = [io.BytesIO() for _ in range(frames + 1)]
fps_avg = []
lag_avg = []

try:
  os.mkdir(dir_name)
except OSError:
  pass

def trigger(next_call, i):
  if i % (hz * 3) == 0:
    print('{}/{} seconds'.format(i/hz, seconds))

  t1_start = time.time()
  camera.capture(imgbuff[i], 'jpeg', use_video_port=True)
  t1_t2 = time.time()
  buff.append(self._tcam.capture())
  t2_stop = time.time()

  sec = t2_stop - t1_start
  fps_avg.append(sec)
  lag_avg.append(t2_stop - t1_t2)

  if sec > (1.0/float(hz)):
    print('Cannot keep up with frame rate!')

  if frames == i:
    return

  th = threading.Timer( next_call - time.time(), trigger,
    args=[next_call+(1.0/float(hz)), i + 1] )
```

```
      th.start()
      th.join()

    trigger(time.time(), 0)

    print('Average time for frame capture = {} seconds'.format(sum(fps_avg)/len(fps_avg)))
    print('Average lag between camera and thermal capture = {} seconds'.format(sum(lag_avg)/len(lag_avg)))

    self.capture_to_file(buff, hz, os.path.join(dir_name, 'output'))

    for i, b in enumerate(imgbuff):
      img_name = os.path.join(dir_name, 'video-{:09d}.jpg'.format(i))
      with open(img_name, 'wb') as f:
        f.write(b.getvalue())

    return (hz, buff)
```

C.1.2 pxdisplay.py

```
from __future__ import division
from __future__ import print_function

from multiprocessing import Process, Queue
import colorsys
import time

def millis_diff(a, b):
  diff = b - a
  return (diff.days * 24 * 60 * 60 + diff.seconds) * 1000 + diff.microseconds / 1000.0

def temp_to_rgb(temp, tmin, tmax):
        OLD_MIN = tmin
```

```python
        OLD_MAX = tmax

        if temp < OLD_MIN:
          temp = OLD_MIN

        if temp > OLD_MAX:
          temp = OLD_MAX

        v = (temp - OLD_MIN) / (OLD_MAX - OLD_MIN)

        rgb = colorsys.hsv_to_rgb((1-v), 1, v * 0.5)

        return tuple(int(c * 255) for c in rgb)

def create(q=None, limit=0, width=100, tmin=15, tmax=45, caption="Display"):
  if q is None:
    q = Queue()

  p = Process(target=_display_process, args=(q, caption, tmin, tmax, limit, width))
  p.daemon = True
  p.start()

  return (q, p)

def _display_process(q, caption, tmin, tmax, limit, pxwidth):
  import pygame
  pygame.init()
  pygame.display.set_caption(caption)

  size = (16 * pxwidth, 4 * pxwidth)
  screen = pygame.display.set_mode(size)

  background = pygame.Surface(screen.get_size())
  background = background.convert_alpha()
```

```
font = pygame.font.Font(None, 36)

while True:
  for event in pygame.event.get():
    if event.type == pygame.QUIT:
      pygame.quit()
      return

  # Keep the event loop running so the windows don't freeze without data
  try:
    qg = q.get(True, 0.3)
  except:
    continue

  px = qg['ir']

  #lag = q.qsize()
  #if lag > 0:
  #  print("WARNING: Dropped " + str(lag) + " frames")

  for i, row in enumerate(px):
    for j, v in enumerate(row):
      rgb = temp_to_rgb(v, tmin, tmax)

      x = i*pxwidth
      y = j*pxwidth

      screen.fill(rgb, (y, x, pxwidth, pxwidth))

  if 'text' in qg:
    background.fill((0, 0, 0, 0))
    text = font.render(qg['text'], 1, (255,255,255))
    background.blit(text, (0,0))
```

```
      # Blit everything to the screen
      screen.blit(background, (0, 0))

    pygame.display.flip()

    if limit != 0:
      time.sleep(1.0/float(limit))
```

C.1.3 features.py

```
from __future__ import division
from __future__ import print_function

import threading
import pxdisplay
import time
import math
import copy
import networkx as nx
import itertools
import collections
#import matplotlib.pyplot as plt

def tuple_to_list(l):
  new = []

  for r in l:
    new.append(list(r))

  return new

```

```python
def min_temps(l, n):
  flat = []
  for i, r in enumerate(l):
    for j, v in enumerate(r):
      flat.append(((i,j), v))
  flat.sort(key=lambda x: x[1])

  ret = [x[0] for x in flat]
  return ret[:n]


def init_arr(val=None):
  return [[val for x in range(16)] for x in range(4)]

class Features(object):
  _q = None
  _thread = None

  _background = None
  _means = None
  _stds = None
  _stds_post = None
  _active = None

  _num_active = None
  _connected_graph = None
  _num_connected = None
  _size_connected = None

  _lock = None

  _rows = None
  _columns = None
```

```
motion_weight = None
nomotion_weight = None

motion_window = None

hz = None

display = None

_exit = False

def __init__(self, q, hz, motion_window=10, motion_weight=0.1, nomotion_weight=0.01, display=True, rows=4, columns=16):
  self._q = q
  self.hz = hz
  self.motion_weight = motion_weight
  self.nomotion_weight = nomotion_weight
  self.display = display
  self.motion_window = motion_window

  self._active = []

  self._rows = rows
  self._columns = columns

  self._thread = threading.Thread(group=None, target=self._monitor_thread)
  self._thread.daemon = True

  self._lock = threading.Lock()

  self._thread.start()

def get_background(self):
  self._lock.acquire()
```

```
    background = copy.deepcopy(self._background)
    self._lock.release()
    return background

  def get_means(self):
    self._lock.acquire()
    means = copy.deepcopy(self._means)
    self._lock.release()
    return means

  def get_stds(self):
    self._lock.acquire()
    stds = copy.deepcopy(self._stds_post)
    self._lock.release()
    return stds

  def get_active(self):
    self._lock.acquire()
    active = copy.deepcopy(self._active)
    self._lock.release()
    return active

  def get_features(self):
    self._lock.acquire()
    num_active = self._num_active
    num_connected = self._num_connected
    size_connected = self._size_connected
    self._lock.release()
    return (num_active, num_connected, size_connected)

  def close(self):
    self._exit = True

    if self._thread is not None:
```

```
      while self._thread.is_alive(): # Wait for thread to terminate
        pass

  def __del__(self):
    self.close()

  def _monitor_thread(self):
    bdisp = None
    ddisp = None

    freq = self.hz * self.motion_window
    mwin = collections.deque([False] * freq)

    n = 1
    while True:
      fdata = None

      if self._exit:
        return

      try:
        fdata = self._q.get(True, 0.3)
      except:
        continue

      if self.display and bdisp is None:
        bdisp, _ = pxdisplay.create(caption="Background", width=80)
        ddisp, _ = pxdisplay.create(caption="Deviation", width=80)

      frame = fdata['ir']

      mwin.popleft()
      mwin.append(fdata['movement'])
      motion = any(mwin)
```

```
self._lock.acquire()

self._active = []

g = nx.Graph()

if n == 1:
  self._background = tuple_to_list(frame)
  self._means = tuple_to_list(frame)
  self._stds = init_arr(0)
  self._stds_post = init_arr()
else:
  weight = self.nomotion_weight
  use_frame = frame

  # Not currently working
  #if motion:
  #  indeces = min_temps(frame, 5)
  #  scalepx = []
  #
  #  for i, j in indeces:
  #    scalepx.append(self._background[i][j] / frame[i][j])
  #
  #  scale = sum(scalepx) / len(scalepx)
  #  scaled_bg = [[x * scale for x in r] for r in frame]
  #
  #  weight = self.motion_weight
  #  use_frame = scaled_bg

  for i in range(self._rows):
    for j in range(self._columns):
      prev = self._background[i][j]
      cur = use_frame[i][j]
```

```
      cur_mean = self._means[i][j]
      cur_std = self._stds[i][j]

      if not motion: # TODO: temp fix
        self._background[i][j]   = weight * cur + (1 - weight) * prev

        # maybe exclude these from motion calculations?
        # n doesn't change when in motion, so it'll cause all sort of corrupted results, as they use n?
        self._means[i][j] = cur_mean + (cur - cur_mean) / n
        self._stds[i][j]  = cur_std + (cur - cur_mean) * (cur - self._means[i][j])
        self._stds_post[i][j] = math.sqrt(self._stds[i][j] / (n-1))

      if (cur - self._background[i][j]) > (3 * self._stds_post[i][j]):
        self._active.append((i,j))

        g.add_node((i,j))

        x = [(-1, -1), (-1, 0), (-1, 1), (0, -1)] # Nodes that have already been computed as active
        for ix, jx in x:
          if (i+ix, j+jx) in self._active:
            g.add_edge((i,j), (i+ix,j+jx))

active = self._active

self._num_active = len(self._active)

components = list(nx.connected_components(g))

self._connected_graph = g
self._num_connected = nx.number_connected_components(g)
self._size_connected = max(len(component) for component in components) if len(components) > 0 else None

self._lock.release()
```

```python

      if self.display:
        bdisp.put({'ir': self._background})

        if n >= 2:
          std = {'ir': init_arr(0)}

          for i, j in active:
            std['ir'][i][j] = frame[i][j]

          ddisp.put(std)

      #print(n)
      #if n > 30:
      #  nx.draw(g)
      #  plt.show()


      if not motion:
        n += 1
```

C.2 Arduino Sketch

```c
/**
 * MLX90260 Arduino Interface
 * Based on code from http://forum.arduino.cc/index.php/topic,126244.0.html
 */
//#define __ASSERT_USE_STDERR

//#include <assert.h>
#include <math.h>
#include <Wire.h>
```

```
#include <EEPROM.h>
#include "SimpleTimer.h" // http://playground.arduino.cc/Code/SimpleTimer

// Configurable options
const int POR_CHECK_FREQ    = 2000; // Time in milliseconds to check if MLX reset has occurred
const int PIR_INTERRUPT_PIN = 0;    // D2 on the Arduino Uno

// Configuration constants
#define PIXEL_LINES     4
#define PIXEL_COLUMNS   16
#define BYTES_PER_PIXEL 2
#define EEPROM_SIZE     255
#define NUM_PIXELS      (PIXEL_LINES * PIXEL_COLUMNS)

// EEPROM helpers
#define E_READ(X)       (EEPROM_DATA[X])
#define E_WRITE(X, Y)   (EEPROM_DATA[X] = (Y))

// Bit fiddling helpers
#define BYTES2INT(H, L)     ( ((H) << 8) + (L) )
#define UBYTES2INT(H, L)    ( ((unsigned int)(H) << 8) + (unsigned int)(L) )
#define BYTE2INT(B)         ( ((int)(B) > 127) ? ((int)(B) - 256) : (int)(B) )
#define E_BYTES2INT(H, L)   ( BYTES2INT(E_READ(H), E_READ(L)) )
#define E_UBYTES2INT(H, L)  ( UBYTES2INT(E_READ(H), E_READ(L)) )
#define E_BYTE2INT(X)       ( BYTE2INT(E_READ(X)) )

// I2C addresses
#define ADDR_EEPROM   0x50
#define ADDR_SENSOR   0x60

// I2C commands
#define CMD_SENSOR_READ         0x02
#define CMD_SENSOR_WRITE_CONF   0x03
#define CMD_SENSOR_WRITE_TRIM   0x04
```

```
// Addresses in the sensor RAM (see Table 9 in spec)
#define SENSOR_PTAT             0x90
#define SENSOR_CPIX             0x91
#define SENSOR_CONFIG           0x92

// Addresses in the EEPROM (see Tables 5 & 7 in spec)
#define EEPROM_A_I_00             0x00 // A_i(0,0) IR pixel individual offset coefficient (ends at 0x3F)
#define EEPROM_B_I_00             0x40 // B_i(0,0) IR pixel individual offset coefficient (ends at 0x7F)
#define EEPROM_DELTA_ALPHA_00     0x80 // Delta-alpha(0,0) IR pixel individual offset coefficient (ends at 0xBF)
#define EEPROM_A_CP               0xD4 // Compensation pixel individual offset coefficients
#define EEPROM_B_CP               0xD5 // Individual Ta dependence (slope) of the compensation pixel offset
#define EEPROM_ALPHA_CP_L         0xD6 // Sensitivity coefficient of the compensation pixel (low)
#define EEPROM_ALPHA_CP_H         0xD7 // Sensitivity coefficient of the compensation pixel (high)
#define EEPROM_TGC                0xD8 // Thermal gradient coefficient
#define EEPROM_B_I_SCALE          0xD9 // Scaling coefficient for slope of IR pixels offset
#define EEPROM_V_TH_L             0xDA // V_TH0 of absolute temperature sensor (low)
#define EEPROM_V_TH_H             0xDB // V_TH0 of absolute temperature sensor (high)
#define EEPROM_K_T1_L             0xDC // K_T1 of absolute temperature sensor (low)
#define EEPROM_K_T1_H             0xDD // K_T1 of absolute temperature sensor (high)
#define EEPROM_K_T2_L             0xDE // K_T2 of absolute temperature sensor (low)
#define EEPROM_K_T2_H             0xDF // K_T2 of absolute temperature sensor (high)
#define EEPROM_ALPHA_0_L          0xE0 // Common sensitivity coefficient of IR pixels (low)
#define EEPROM_ALPHA_0_H          0xE1 // Common sensitivity coefficient of IR pixels (high)
#define EEPROM_ALPHA_0_SCALE      0xE2 // Scaling coefficient for common sensitivity
#define EEPROM_DELTA_ALPHA_SCALE  0xE3 // Scaling coefficient for individual sensitivity
#define EEPROM_EPSILON_L          0xE4 // Emissivity (low)
#define EEPROM_EPSILON_H          0xE5 // Emissivity (high)
#define EEPROM_TRIMMING_VAL       0xF7 // Oscillator trimming value

// Config flag locations
#define CFG_TA    8
#define CFG_IR    9
#define CFG_POR   10
```

```
// Arduino EEPROM addresses
#define AEEP_FREQ_ADDR 0x00

// Global variables
unsigned int PTAT;              // Proportional to absolute temperature value
int CPIX;                       // Compensation pixel

int IRDATA[NUM_PIXELS];         // Infrared raw data
byte EEPROM_DATA[EEPROM_SIZE];  // EEPROM dump

float ta;                       // Absolute chip temperature / ambient chip temperature (degrees celsius)
float emissivity;               // Emissivity compensation
float k_t1;                     // K_T1 of absolute temperature sensor
float k_t2;                     // K_T2 of absolute temperature sensor
float da0_scale;                // Scaling coefficient for individual sensitivity
float alpha_const;              // Common sensitivity coefficient of IR pixels and scaling coefficient for common sensitivity

int v_th;                       // V_TH0 of absolute temperature sensor
int a_cp;                       // Compensation pixel individual offset coefficients
int b_cp;                       // Individual Ta dependence (slope) of the compensation pixel offset
int tgc;                        // Thermal gradient coefficient
int b_i_scale;                  // Scaling coefficient for slope of IR pixels offset

float alpha_ij[NUM_PIXELS];     // Individual pixel sensitivity coefficient
int a_ij[NUM_PIXELS];           // Individual pixel offset
int b_ij[NUM_PIXELS];           // Individual pixel offset slope coefficient

char hpbuf[2];                  // Hex printing buffer
int res;                        // Error code storage

float temp[NUM_PIXELS];         // Final calculated temperature values in degrees celsius
```

```
SimpleTimer timer;              // Allows timed callbacks for temp functions

void(* reset_arduino_now) (void) = 0;   // Creates function to reset Arduino

// Stores references to the 3 timers used in the program
int ir_timer;
int ta_timer;
int por_timer;

// Stores refresh frequency, read out of the EEPROM
short REFRESH_FREQ;

volatile bool pir_motion_detected = false;

/*
// Send assertion failures over serial
void __assert(const char *__func, const char *__file, int __lineno, const char *__sexp) {
    // transmit diagnostic informations through serial link.
    Serial.println(__func);
    Serial.println(__file);
    Serial.println(__lineno, DEC);
    Serial.println(__sexp);
    Serial.flush();
    // abort program execution.
    abort();
}*/

void reset_arduino() {
  Serial.flush();
  reset_arduino_now();
}

// Basic assertion failure function
void assert(boolean a) {
```

```
  if (!a) Serial.println("ASSFAIL");
}

// Takes byte value and will output 2 character hex representation on serial
void print_hex(byte b) {
  hpbuf[0] = (b >> 4) + 0x30;
  if (hpbuf[0] > 0x39) hpbuf[0] +=7;

  hpbuf[1] = (b & 0x0f) + 0x30;
  if (hpbuf[1] > 0x39) hpbuf[1] +=7;

  Serial.print(hpbuf);
}

// Will read memory from the given sensor address and convert it into an integer
int _sensor_read_int(byte read_addr) {
  Wire.beginTransmission(ADDR_SENSOR);
  Wire.write(CMD_SENSOR_READ);
  Wire.write(read_addr);
  Wire.write(0x00); // address step (0)
  Wire.write(0x01); // number of reads (1)
  res = Wire.endTransmission(false); // we must use the repeated start here
  if (res != 0) return -1;

  Wire.requestFrom(ADDR_SENSOR, 2); // technically the 1 read takes up 2 bytes

  int LSB, MSB;
  int i = 0;
  while( Wire.available() ) {
    i++;

    if (i > 2) {
      return -1; // Returned more bytes than it should have
    }
```

```
    LSB = Wire.read();
    MSB = Wire.read();
  }

  return UBYTES2INT(MSB, LSB); // rearrange int to account for endian difference (TODO: check)
}

// Will read a configuration flag bit specified by flag_loc from the sensor config
bool _sensor_read_config_flag(int flag_loc) {
  int cur_cfg = _sensor_read_int(SENSOR_CONFIG);
  return (bool)(cur_cfg & ( 1 << flag_loc )) >> flag_loc;
}

// Reads Proportional To Absolute Temperature (PTAT) value
int sensor_read_ptat() {
  return _sensor_read_int(SENSOR_PTAT);
}

// Reads compensation pixel
int sensor_read_cpix() {
  return _sensor_read_int(SENSOR_CPIX);
}

// Reads POR flag
bool sensor_read_por() {
  return _sensor_read_config_flag(CFG_POR); // POR is 10th bit
}

// Read Ta measurement flag
bool sensor_read_ta_measure() {
  return _sensor_read_config_flag(CFG_TA);
}
```

```
// Read IR measurement flag
bool sensor_read_ir_measure() {
  return _sensor_read_config_flag(CFG_IR);
}

// Reads all raw IR data from sensor into IRDATA variable
boolean sensor_read_irdata() {
  int i = 0;

  // Due to wire library buffer limitations, we can only read up to 32 bytes at a time
  // Thus, the request has been split into multiple different requests to get the full 128 values
  // Each pixel value takes up two bytes (???) thus NUM_PIXELS * 2
  for (int line = 0; line < PIXEL_LINES; line++) {
    Wire.beginTransmission(ADDR_SENSOR);
    Wire.write(CMD_SENSOR_READ);
    Wire.write(line);
    Wire.write(0x04);
    Wire.write(0x10);
    res = Wire.endTransmission(false); // use repeated start to get answer

    if (res != 0) return false;

    Wire.requestFrom(ADDR_SENSOR, PIXEL_COLUMNS * BYTES_PER_PIXEL);

    byte PIX_LSB, PIX_MSB;

    for(int j = 0; j < PIXEL_COLUMNS; j++) {
      if (!Wire.available()) return false;

      // We read two bytes
      PIX_LSB = Wire.read();
      PIX_MSB = Wire.read();

      IRDATA[i] = BYTES2INT(PIX_MSB, PIX_LSB);
```

```
      i++;
    }
  }

  return true;
}

// Will send a command and the provided most significant and least significant bit
// with the appropriate check bit added
// Returns the Wire success/error code
boolean _sensor_write_check(byte cmd, byte check, byte lsb, byte msb) {
  Wire.beginTransmission(ADDR_SENSOR);
  Wire.write(cmd);            // Send the command
  Wire.write(lsb - check);  // Send the least significant byte check
  Wire.write(lsb);            // Send the least significant byte
  Wire.write(msb - check);  // Send the most significant byte check
  Wire.write(msb);            // Send the most significant byte
  return Wire.endTransmission() == 0;
}

// See datasheet: 9.4.2 Write configuration register command
// See datasheet: 8.2.2.1 Configuration register (0x92)
// Check byte is 0x55 in this instance
boolean sensor_write_conf() {
  byte cfg_MSB = B01110100;
  //              ||||||||
  //              |||||||*--- Ta measurement running (read only)
  //              ||||||*---- IR measurement running (read only)
  //              |||||*----- POR flag cleared
  //              ||||*------ I2C FM+ mode enabled
  //              ||**------- Ta refresh rate (2 byte code, 2Hz hardcoded)
  //              |*--------- ADC high reference
  //              *---------- NA

```

```
byte cfg_LSB = B00001110;
//              ||||||||
//              ||||****--- 4 byte IR refresh rate (4 byte code, 1Hz default)
//              ||**------- NA
//              |*--------- Continuous measurement mode
//              *---------- Normal operation mode

switch(REFRESH_FREQ) {
case 0: // 0.5Hz
  cfg_LSB = B00001111;
  break;
case 2:
  cfg_LSB = B00001101;
  break;
case 4:
  cfg_LSB = B00001100;
  break;
case 8:
  cfg_LSB = B00001011;
  break;
case 16:
  cfg_LSB = B00001010;
  break;
case 32:
  cfg_LSB = B00001001;
  break;
case 64:
  cfg_LSB = B00001000;
  break;
case 128:
  cfg_LSB = B00000111;
  break;
case 256:
  cfg_LSB = B00000110;
```

```
    break;
  case 512:
    cfg_LSB = B00000000; // modes 5 to 0 are all 512Hz
    break;
  }

  return _sensor_write_check(CMD_SENSOR_WRITE_CONF, 0x55, cfg_LSB, cfg_MSB);
}

// See datasheet: 9.4.3 Write trimming command
// Check byte is 0xAA in this instance
boolean sensor_write_trim() {
  return _sensor_write_check(CMD_SENSOR_WRITE_TRIM, 0xAA, E_READ(EEPROM_TRIMMING_VAL), 0x00);
}

// Reads EEPROM memory into global variable
boolean eeprom_read_all() {
  int i = 0;
  // Due to wire library buffer limitations, we can only read up to 32 bytes at a time
  // Thus, the request has been split into 4 different requests to get the full 128 values
  for(int j = 0; j < EEPROM_SIZE; j = j + 32) {
    Wire.beginTransmission(ADDR_EEPROM);
    Wire.write( byte(j) );
    res = Wire.endTransmission();

    if (res != 0) return false;

    Wire.requestFrom(ADDR_EEPROM, 32);

    i = j;
    while( Wire.available() ) { // slave may send less than requested
      byte b = Wire.read(); // receive a byte as character
      E_WRITE(i, b);
      i++;
```

```
    }
  }

  if (i < EEPROM_SIZE) { // If we didn't get the whole EEPROM
    return false;
  }

  return true;
}

// Writes various calculation values from EEPROM into global variables
void calculate_init() {
  v_th = E_BYTES2INT(EEPROM_V_TH_H, EEPROM_V_TH_L);
  k_t1 = E_BYTES2INT(EEPROM_K_T1_H, EEPROM_K_T1_L) / 1024.0;
  k_t2 = E_BYTES2INT(EEPROM_K_T2_H, EEPROM_K_T2_L) / 1048576.0;

  a_cp = E_BYTE2INT(EEPROM_A_CP);
  b_cp = E_BYTE2INT(EEPROM_B_CP);
  tgc  = E_BYTE2INT(EEPROM_TGC);

  b_i_scale = E_READ(EEPROM_B_I_SCALE);

  emissivity = E_UBYTES2INT(EEPROM_EPSILON_H, EEPROM_EPSILON_L) / 32768.0;

  da0_scale = pow(2, -E_READ(EEPROM_DELTA_ALPHA_SCALE));
  alpha_const = (float)E_UBYTES2INT(EEPROM_ALPHA_0_H, EEPROM_ALPHA_0_L) * pow(2, -E_READ(EEPROM_ALPHA_0_SCALE));

  for (int i = 0; i < NUM_PIXELS; i++){
    float alpha_var = (float)E_READ(EEPROM_DELTA_ALPHA_00 + i) * da0_scale;
    alpha_ij[i] = (alpha_const + alpha_var);

    a_ij[i] = E_BYTE2INT(EEPROM_A_I_00 + i);
    b_ij[i] = E_BYTE2INT(EEPROM_B_I_00 + i);
  }
```

```
}

// Calculates the absolute chip temperature from the proportional to absolute temperature (PTAT)
float calculate_ta() {
  float ptat = (float)sensor_read_ptat();
  assert(ptat != -1);
  return (-k_t1 +
      sqrt(
        square(k_t1) -
        ( 4 * k_t2 * (v_th-ptat) )
      )
    ) / (2*k_t2) + 25;
}

// Calculates the final temperature value for each pixel and stores it in temp array
void calculate_temp() {
  float v_cp_off_comp = (float) CPIX - (a_cp + (b_cp/pow(2, b_i_scale)) * (ta - 25));

  for (int i = 0; i < NUM_PIXELS; i++){
    float alpha_ij_v = alpha_ij[i];
    int a_ij_v = a_ij[i];
    int b_ij_v = b_ij[i];

    float v_ir_tgc_comp = IRDATA[i] - (a_ij_v + (float)(b_ij_v/pow(2, b_i_scale)) * (ta - 25)) -
     ↪  (((float)tgc/32)*v_cp_off_comp);
    float v_ir_comp = v_ir_tgc_comp / emissivity;
    temp[i] = sqrt(sqrt((v_ir_comp/alpha_ij_v) + pow((ta + 273.15),4))) - 273.15;
  }

}

// Prints all of EEPROM as hex
void print_eeprom() {
  Serial.print("EEPROM ");
```

```
  for(int i = 0; i < EEPROM_SIZE; i++) {
    print_hex(E_READ(i));
  }
  Serial.println();
}

// Prints a serial "packet" containing IR data
void print_packet(unsigned long cur_time) {
  Serial.print("START ");
  Serial.println(cur_time);

  Serial.print("MOVEMENT ");
  Serial.println(pir_motion_detected);

  for(int i = 0; i<NUM_PIXELS; i++) {
    Serial.print(temp[i]);

    if ((i+1) % PIXEL_COLUMNS == 0) {
      Serial.println();
    } else {
      Serial.print("\t");
    }
  }

 Serial.print("STOP ");
 Serial.println(millis());
 Serial.flush();
}

// Prints info about driver, build and configuration
void print_info() {
  Serial.println("INFO START");
  Serial.println("DRIVER MLX90620");

```

```
  Serial.print("BUILD ");
  Serial.print(__DATE__);
  Serial.print(" ");
  Serial.println(__TIME__);

  Serial.print("IRHZ ");
  Serial.println(REFRESH_FREQ);
  Serial.println("INFO STOP");
}

// Runs functions necessary to initialize the temperature sensor
void initialize() {
  assert(eeprom_read_all());
  assert(sensor_write_trim());
  assert(sensor_write_conf());

  calculate_init();

  ta_loop();
}

// Calculates absolute temperature
void ta_loop() {
  ta = calculate_ta();
}

// Checks if the sensor as been reset, and if so, re-runs the initialize functions
void por_loop() {
  if (!sensor_read_por()) { // there has been a reset
    initialize();
  }
}

// Runs functions necessary to compute and output the temperature data
```

```
void ir_loop() {
  unsigned long cur_time = millis();

  assert(sensor_read_irdata());

  CPIX = sensor_read_cpix();
  assert(CPIX != -1);

  calculate_temp();

  print_packet(cur_time);

  pir_motion_detected = false;
}

// Configures timers to poll IR and other data periodically
void activate_timers() {
  float hz = REFRESH_FREQ;

  if (REFRESH_FREQ == 0) {
    hz = 0.5;
  }

  // Calculate how many milliseconds each timer should run for
  // based upon the configured refresh rate of the IR data and
  // absolute temperature data
  long irlen = (1/hz) * 1000;
  long talen = (1/2.0) * 1000;

  if (talen < irlen) {
    talen = irlen;
  }

  ir_timer = timer.setInterval(irlen, ir_loop);
```

```
  ta_timer = timer.setInterval(talen, ta_loop);
  por_timer = timer.setInterval(POR_CHECK_FREQ, por_loop);

  attachInterrupt(PIR_INTERRUPT_PIN, pir_motion, RISING);
}

// Disables timers to poll IR and other data periodically
void deactivate_timers() {
  timer.disable(ir_timer);
  timer.deleteTimer(ir_timer);

  timer.disable(ta_timer);
  timer.deleteTimer(ta_timer);

  timer.disable(por_timer);
  timer.deleteTimer(por_timer);

  detachInterrupt(PIR_INTERRUPT_PIN);
}

void pir_motion() {
  pir_motion_detected = true;
}

void read_freq() {
  byte rd = EEPROM.read(0);

  if (rd > 9) {
    rd = 0;
    EEPROM.write(AEEP_FREQ_ADDR, 0);
  }

  switch(rd) {
  case 1:
```

```
  REFRESH_FREQ = 1;
  break;
case 2:
  REFRESH_FREQ = 2;
  break;
case 3:
  REFRESH_FREQ = 4;
  break;
case 4:
  REFRESH_FREQ = 8;
  break;
case 5:
  REFRESH_FREQ = 16;
  break;
case 6:
  REFRESH_FREQ = 32;
  break;
case 7:
  REFRESH_FREQ = 64;
  break;
case 8:
  REFRESH_FREQ = 128;
  break;
case 9:
  REFRESH_FREQ = 256;
  break;
case 10:
  REFRESH_FREQ = 512;
  break;

default:
case 0:
  REFRESH_FREQ = 0;
  break;
```

```
  }
}

void write_freq(int freq) {
  byte wt;

  switch(freq) {
  case 1:
    wt = 1;
    break;
  case 2:
    wt = 2;
    break;
  case 4:
    wt = 3;
    break;
  case 8:
    wt = 4;
    break;
  case 16:
    wt = 5;
    break;
  case 32:
    wt = 6;
    break;
  case 64:
    wt = 7;
    break;
  case 128:
    wt = 8;
    break;
  case 256:
    wt = 9;
    break;
```

```
  case 512: // writing 512 to the config doesn't work for some reason
    wt = 10;
    break;

  default:
  case 0:
    wt = 0;
    break;
  }

  EEPROM.write(AEEP_FREQ_ADDR, wt);
}

// Configure libraries and sensors at startup
void setup() {
  pinMode(2, INPUT);

  Wire.begin();
  Serial.begin(115200);

  Serial.println();
  Serial.print("INIT ");
  Serial.println(millis());

  read_freq();
  print_info();
  initialize();

  Serial.print("ACTIVE ");
  Serial.println(millis());
  Serial.flush();
}

char manualLoop = 0;
```

```
// Triggered when serial data is sent to Arduino. Used to trigger basic actions.
void serialEvent() {
  while (Serial.available()) {
    char in = (char)Serial.read();
    if (in == '\r' || in == '\n') continue;

    switch (in) {
    case 'R':
    case 'r':
      reset_arduino();
      break;

    case 'I':
    case 'i':
      print_info();
      break;

    case 'T':
    case 't':
      activate_timers();
      break;

    case 'O':
    case 'o':
      deactivate_timers();
      break;

    case 'P':
    case 'p':
      if (manualLoop == 16) { // Run ta_loop every 16 manual iterations
        ta_loop();
        manualLoop = 0;
      }
```

```
      ir_loop();

      manualLoop++;
      break;

    case 'f':
    case 'F':
      write_freq(Serial.parseInt());
      reset_arduino();
      break;

    default:
      Serial.println("UNKNOWN COMMAND");
    }
  }
}

void loop() {
  timer.run();
}
```

APPENDIX D

Full Results

D.1 Classifier Experiment Set 1

D.1.1 combined-exp-all-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filte


Correctly Classified Instances         250               24.6063 %
Incorrectly Classified Instances       766               75.3937 %
Kappa statistic                          -0.0052
Mean absolute error                       0.375
Root mean squared error                   0.433
Relative absolute error                 100      %
Root relative squared error             100      %
Coverage of cases (0.95 level)          100      %
Mean rel. region size (0.95 level)      100      %
Total Number of Instances              1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.591    0.604    0.246      0.591    0.347      -0.012
                  ↪  0.493    0.247    0
                 0.394    0.402    0.246      0.394    0.303      -0.007
                  ↪  0.494    0.248    1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.494    0.248    2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.493    0.247    3
```

```
Weighted Avg.    0.246    0.251    0.123      0.246    0.163      -0.005
 ↪  0.493     0.247

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 150 104   0   0 |   a = 0
 154 100   0   0 |   b = 1
 152 102   0   0 |   c = 2
 154 100   0   0 |   d = 3
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances        625               61.5157 %
Incorrectly Classified Instances      391               38.4843 %
Kappa statistic                         0.4869
Mean absolute error                     0.2317
Root mean squared error                 0.3678
Relative absolute error                61.774  %
Root relative squared error            84.9349 %
Coverage of cases (0.95 level)         93.5039 %
Mean rel. region size (0.95 level)     58.0463 %
Total Number of Instances            1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.831    0.136    0.670      0.831    0.742      0.650
                  ↪  0.899    0.712     0
                 0.441    0.146    0.502      0.441    0.470      0.309
                  ↪  0.783    0.559     1
                 0.461    0.157    0.494      0.461    0.477      0.310
                  ↪  0.755    0.454     2
                 0.728    0.073    0.768      0.728    0.747      0.667
                  ↪  0.920    0.793     3
Weighted Avg.    0.615    0.128    0.608      0.615    0.609      0.484
 ↪  0.839     0.629

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 211  32  10   1 |   a = 0
  85 112  52   5 |   b = 1
  19  68 117  50 |   c = 2
```

```
   0  11  58 185 |   d = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.781
Mean absolute error                      0.4593
Root mean squared error                  0.664
Relative absolute error                 54.9126 %
Root relative squared error             68.4298 %
Total Number of Instances              305
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.1293
Mean absolute error                      0.8203
Root mean squared error                  0.9719
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correlation coefficient                  0.7341
Mean absolute error                      0.5098
Root mean squared error                  0.6587
Relative absolute error                 62.1465 %
Root relative squared error             67.7669 %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter
```

```
Correctly Classified Instances         667               65.6496 %
Incorrectly Classified Instances       349               34.3504 %
Kappa statistic                          0.542
Mean absolute error                      0.1915
Root mean squared error                  0.3643
Relative absolute error                 51.0556 %
Root relative squared error             84.1382 %
Coverage of cases (0.95 level)          82.7756 %
Mean rel. region size (0.95 level)      41.437  %
Total Number of Instances             1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.358    0.035    0.771      0.358    0.489      0.436
                  ↪  0.866     0.688     0
                 0.638    0.046    0.822      0.638    0.718      0.648
                  ↪  0.917     0.801     1
                 0.846    0.087    0.765      0.846    0.804      0.736
                  ↪  0.933     0.859     2
                 0.783    0.290    0.474      0.783    0.591      0.434
                  ↪  0.861     0.649     3
Weighted Avg.    0.656    0.115    0.708      0.656    0.650      0.564
 ↪  0.894     0.749

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
  91  15   0 148 |   a = 0
  26 162  18  48 |   b = 1
   1  13 215  25 |   c = 2
   0   7  48 199 |   d = 3
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.3766
Mean absolute error                      0.7113
Root mean squared error                  1.1235
Relative absolute error                 86.7178 %
Root relative squared error            115.5897 %
```

```
Total Number of Instances            1018
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         578               56.8898 %
Incorrectly Classified Instances       438               43.1102 %
Kappa statistic                          0.4252
Mean absolute error                      0.3014
Root mean squared error                  0.3855
Relative absolute error                 80.3795 %
Root relative squared error             89.0256 %
Coverage of cases (0.95 level)          97.1457 %
Mean rel. region size (0.95 level)      75.0738 %
Total Number of Instances             1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.732    0.189    0.564      0.732    0.637      0.502
                  ↪  0.818     0.517     0
                 0.378    0.186    0.403      0.378    0.390      0.196
                  ↪  0.697     0.364     1
                 0.409    0.098    0.581      0.409    0.480      0.354
                  ↪  0.739     0.435     2
                 0.756    0.101    0.714      0.756    0.734      0.643
                  ↪  0.889     0.657     3
Weighted Avg.    0.569    0.144    0.565      0.569    0.560      0.424
 ↪  0.786     0.493

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 186  64   3   1 |   a = 0
 124  96  26   8 |   b = 1
  20  62 104  68 |   c = 2
   0  16  46 192 |   d = 3
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter
```

```
Correctly Classified Instances         788               77.5591 %
Incorrectly Classified Instances       228               22.4409 %
Kappa statistic                          0.7008
Mean absolute error                      0.167
Root mean squared error                  0.2898
Relative absolute error                 44.5337 %
Root relative squared error             66.9142 %
Coverage of cases (0.95 level)          96.4567 %
Mean rel. region size (0.95 level)      57.5787 %
Total Number of Instances             1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.776    0.087    0.749      0.776    0.762      0.681
                  ↪  0.931     0.707     0
                 0.681    0.088    0.721      0.681    0.700      0.605
                  ↪  0.888     0.727     1
                 0.823    0.084    0.766      0.823    0.793      0.722
                  ↪  0.896     0.812     2
                 0.823    0.041    0.871      0.823    0.846      0.797
                  ↪  0.956     0.867     3
Weighted Avg.    0.776    0.075    0.777      0.776    0.775      0.701
 ↪  0.918     0.778

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 197  39  14   4 |   a = 0
  58 173  13  10 |   b = 1
   8  20 209  17 |   c = 2
   0   8  37 209 |   d = 3
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         220               72.1311 %
Incorrectly Classified Instances        85               27.8689 %
Kappa statistic                          0.6273
Mean absolute error                      0.1868
Root mean squared error                  0.3197
Relative absolute error                 49.7617 %
```

```
Root relative squared error             73.7524 %
Coverage of cases (0.95 level)          93.1148 %
Mean rel. region size (0.95 level)      50.4918 %
Total Number of Instances              305

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.870    0.105    0.736      0.870    0.798      0.726
                  ↪  0.917     0.699     0
                 0.698    0.114    0.706      0.698    0.702      0.586
                  ↪  0.870     0.748     1
                 0.681    0.081    0.712      0.681    0.696      0.610
                  ↪  0.892     0.677     2
                 0.630    0.073    0.730      0.630    0.676      0.587
                  ↪  0.903     0.785     3
Weighted Avg.    0.721    0.095    0.721      0.721    0.719      0.627
 ↪  0.894     0.728

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 67 10  0  0 |  a = 0
 21 60  2  3 |  b = 1
  3  5 47 14 |  c = 2
  0 10 17 46 |  d = 3
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.828
Mean absolute error                      0.3829
Root mean squared error                  0.5496
Relative absolute error                 46.674  %
Root relative squared error             56.542  %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filte
```

```
Correctly Classified Instances         795               78.248  %
Incorrectly Classified Instances       221               21.752  %
Kappa statistic                          0.71
Mean absolute error                      0.1898
Root mean squared error                  0.2932
Relative absolute error                 50.6237 %
Root relative squared error             67.7026 %
Coverage of cases (0.95 level)          98.622  %
Mean rel. region size (0.95 level)      64.0256 %
Total Number of Instances             1016

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.870    0.088    0.767      0.870    0.815      0.751
                  ↪  0.941     0.784     0
                 0.697    0.085    0.731      0.697    0.714      0.622
                  ↪  0.914     0.802     1
                 0.811    0.087    0.757      0.811    0.783      0.708
                  ↪  0.912     0.797     2
                 0.752    0.030    0.893      0.752    0.816      0.766
                  ↪  0.962     0.899     3
Weighted Avg.    0.782    0.073    0.787      0.782    0.782      0.712
 ↪  0.932     0.820

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 221  32   0   1 |   a = 0
  59 177  13   5 |   b = 1
   8  23 206  17 |   c = 2
   0  10  53 191 |   d = 3
```

D.1.2 combined-exp-all-noresample-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filte


Correctly Classified Instances         411               40.3733 %
Incorrectly Classified Instances       607               59.6267 %
Kappa statistic                          0
Mean absolute error                      0.3452
Root mean squared error                  0.4154
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.404      1.000    0.575      0.000
                  ↪  0.497     0.402     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.495     0.246     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.497     0.285     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.480     0.061     3
Weighted Avg.    0.404    0.404    0.163      0.404    0.232      0.000
 ↪  0.496     0.308

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 411   0   0   0 |   a = 0
 252   0   0   0 |   b = 1
 291   0   0   0 |   c = 2
  64   0   0   0 |   d = 3
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filte
```

```
Correctly Classified Instances         674               66.2083 %
Incorrectly Classified Instances       344               33.7917 %
Kappa statistic                          0.4964
Mean absolute error                      0.2087
Root mean squared error                  0.3516
Relative absolute error                 60.4564 %
Root relative squared error             84.6405 %
Coverage of cases (0.95 level)          93.0255 %
Mean rel. region size (0.95 level)      58.9882 %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.925    0.209    0.750      0.925    0.828      0.702
                  ↪  0.889     0.808     0
                 0.357    0.127    0.481      0.357    0.410      0.257
                  ↪  0.746     0.523     1
                 0.608    0.149    0.621      0.608    0.615      0.463
                  ↪  0.826     0.683     2
                 0.422    0.013    0.692      0.422    0.524      0.518
                  ↪  0.914     0.515     3
Weighted Avg.    0.662    0.159    0.643      0.662    0.644      0.512
 ↪  0.837     0.683

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 380  16  15   0 |   a = 0
 100  90  60   2 |   b = 1
  27  77 177  10 |   c = 2
   0   4  33  27 |   d = 3
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.1293
Mean absolute error                      0.8203
Root mean squared error                  0.9719
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
```

```
Options: -S 0 -R 1.0E-8
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correlation coefficient                  0.7341
Mean absolute error                      0.5098
Root mean squared error                  0.6587
Relative absolute error                 62.1465 %
Root relative squared error             67.7669 %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.3766
Mean absolute error                      0.7113
Root mean squared error                  1.1235
Relative absolute error                 86.7178 %
Root relative squared error            115.5897 %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         598               58.7426 %
Incorrectly Classified Instances       420               41.2574 %
Kappa statistic                          0.4437
Mean absolute error                      0.2294
Root mean squared error                  0.4117
Relative absolute error                 66.4475 %
Root relative squared error             99.1128 %
Coverage of cases (0.95 level)          70.5305 %
Mean rel. region size (0.95 level)      48.723  %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.287    0.046    0.808      0.287    0.424      0.337
                  ↪  0.870     0.779     0
                 0.802    0.309    0.460      0.802    0.585      0.429
                  ↪  0.833     0.655     1
                 0.859    0.077    0.817      0.859    0.838      0.771
                  ↪  0.928     0.849     2
                 0.438    0.104    0.220      0.438    0.293      0.245
                  ↪  0.796     0.311     3
Weighted Avg.    0.587    0.124    0.688      0.587    0.574      0.478
 ↪  0.873     0.739

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 118 199   6  88 |   a = 0
  27 202  21   2 |   b = 1
   1  31 250   9 |   c = 2
   0   7  29  28 |   d = 3
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         585               57.4656 %
Incorrectly Classified Instances       433               42.5344 %
Kappa statistic                          0.3603
Mean absolute error                      0.297
Root mean squared error                  0.3795
Relative absolute error                 86.0431 %
Root relative squared error             91.3677 %
Coverage of cases (0.95 level)          98.8212 %
Mean rel. region size (0.95 level)      75      %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.740    0.338    0.597      0.740    0.661      0.394
                  ↪  0.777     0.605     0
                 0.353    0.167    0.410      0.353    0.380      0.196
                  ↪  0.625     0.318     1
```

```
                 0.649    0.132    0.663      0.649    0.656      0.521
                  ↪  0.774     0.551     2
                 0.047    0.004    0.429      0.047    0.085      0.125
                  ↪  0.855     0.292     3
Weighted Avg.    0.575    0.216    0.559      0.575    0.554      0.365
 ↪  0.743     0.499

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 304 101   6   0 |   a = 0
 128  89  35   0 |   b = 1
  76  22 189   4 |   c = 2
   1   5  55   3 |   d = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.781
Mean absolute error                      0.4593
Root mean squared error                  0.664
Relative absolute error                 54.9126 %
Root relative squared error             68.4298 %
Total Number of Instances              305
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         844               82.9077 %
Incorrectly Classified Instances       174               17.0923 %
Kappa statistic                          0.7487
Mean absolute error                      0.1731
Root mean squared error                  0.2878
Relative absolute error                 50.1407 %
Root relative squared error             69.3014 %
Coverage of cases (0.95 level)          96.7583 %
Mean rel. region size (0.95 level)      67.5098 %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.929    0.102    0.860      0.929    0.894      0.819
                  ↪  0.925     0.837     0
                 0.710    0.063    0.789      0.710    0.747      0.672
                  ↪  0.855     0.714     1
                 0.873    0.073    0.827      0.873    0.849      0.788
                  ↪  0.910     0.774     2
                 0.453    0.012    0.725      0.453    0.558      0.552
                  ↪  0.870     0.516     3
Weighted Avg.    0.829    0.078    0.825      0.829    0.824      0.756
 ↪  0.900     0.768

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 382  23   5   1 |   a = 0
  52 179  19   2 |   b = 1
  10  19 254   8 |   c = 2
   0   6  29  29 |   d = 3
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.828
Mean absolute error                      0.3829
Root mean squared error                  0.5496
Relative absolute error                 46.674  %
Root relative squared error             56.542  %
Total Number of Instances             1018
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         841               82.613  %
Incorrectly Classified Instances       177               17.387  %
Kappa statistic                          0.7441
Mean absolute error                      0.1764
Root mean squared error                  0.2853
Relative absolute error                 51.1162 %
Root relative squared error             68.6796 %
```

```
Coverage of cases (0.95 level)          98.3301 %
Mean rel. region size (0.95 level)      65.668  %
Total Number of Instances             1018

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.915    0.089    0.874      0.915    0.894      0.820
                  ↪  0.935     0.871     0
                 0.746    0.072    0.774      0.746    0.760      0.683
                  ↪  0.870     0.770     1
                 0.880    0.085    0.805      0.880    0.841      0.774
                  ↪  0.923     0.819     2
                 0.328    0.006    0.778      0.328    0.462      0.486
                  ↪  0.923     0.528     3
Weighted Avg.    0.826    0.078    0.824      0.826    0.818      0.752
 ↪  0.915     0.810

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 376  30   5   0 |   a = 0
  44 188  20   0 |   b = 1
  10  19 256   6 |   c = 2
   0   6  37  21 |   d = 3
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filter


Correctly Classified Instances         240               78.6885 %
Incorrectly Classified Instances        65               21.3115 %
Kappa statistic                          0.6795
Mean absolute error                      0.1425
Root mean squared error                  0.286
Relative absolute error                 41.4629 %
Root relative squared error             69.3884 %
Coverage of cases (0.95 level)          90.1639 %
Mean rel. region size (0.95 level)      48.2787 %
Total Number of Instances              305

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
```

```
                 0.910    0.140    0.834      0.910    0.871      0.765
                  ↪  0.930     0.881     0
                 0.614    0.072    0.717      0.614    0.662      0.573
                  ↪  0.876     0.682     1
                 0.835    0.077    0.807      0.835    0.821      0.750
                  ↪  0.910     0.841     2
                 0.294    0.024    0.417      0.294    0.345      0.318
                  ↪  0.825     0.402     3
Weighted Avg.    0.787    0.100    0.776      0.787    0.779      0.692
 ↪  0.906     0.797

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 121   9   3   0 |   a = 0
  19  43   8   0 |   b = 1
   3   4  71   7 |   c = 2
   2   4   6   5 |   d = 3
```

D.1.3 combined-exp-excl0-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         190               32.4786 %
Incorrectly Classified Instances       395               67.5214 %
Kappa statistic                         -0.0128
Mean absolute error                      0.4445
Root mean squared error                  0.4714
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.487    0.513    0.322      0.487    0.388      -0.024
                  ↪  0.487     0.328     1
                 0.487    0.500    0.328      0.487    0.392      -0.012
                  ↪  0.494     0.331     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.494     0.331     3
Weighted Avg.    0.325    0.338    0.217      0.325    0.260      -0.012
 ↪  0.491     0.330

=== Confusion Matrix ===

   a   b   c   <-- classified as
  95 100   0 |   a = 1
 100  95   0 |   b = 2
 100  95   0 |   c = 3
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         373               63.7607 %
Incorrectly Classified Instances       212               36.2393 %
Kappa statistic                          0.4564
```

```
Mean absolute error                      0.2858
Root mean squared error                  0.4093
Relative absolute error                 64.3009 %
Root relative squared error             86.8329 %
Coverage of cases (0.95 level)          94.359  %
Mean rel. region size (0.95 level)      70.2564 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.800    0.221    0.645      0.800    0.714      0.555
                  ↪  0.866     0.696     1
                 0.374    0.182    0.507      0.374    0.431      0.210
                  ↪  0.685     0.473     2
                 0.738    0.141    0.724      0.738    0.731      0.594
                  ↪  0.878     0.776     3
Weighted Avg.    0.638    0.181    0.625      0.638    0.625      0.453
 ↪  0.810     0.648

=== Confusion Matrix ===

   a   b   c   <-- classified as
 156  34   5 |   a = 1
  72  73  50 |   b = 2
  14  37 144 |   c = 3
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         462               78.9744 %
Incorrectly Classified Instances       123               21.0256 %
Kappa statistic                          0.6846
Mean absolute error                      0.1539
Root mean squared error                  0.3049
Relative absolute error                 34.6168 %
Root relative squared error             64.6801 %
Coverage of cases (0.95 level)          95.5556 %
Mean rel. region size (0.95 level)      51.2821 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.846    0.056    0.882      0.846    0.864      0.798
                  ↪  0.965     0.911     1
                 0.779    0.156    0.714      0.779    0.745      0.610
                  ↪  0.897     0.827     2
                 0.744    0.103    0.784      0.744    0.763      0.650
                  ↪  0.931     0.875     3
Weighted Avg.    0.790    0.105    0.793      0.790    0.791      0.686
 ↪  0.931     0.871

=== Confusion Matrix ===

   a   b   c   <-- classified as
 165  16  14 |   a = 1
  17 152  26 |   b = 2
   5  45 145 |   c = 3
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.1184
Mean absolute error                      0.5641
Root mean squared error                  0.6507
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              585
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correlation coefficient                  0.5887
Mean absolute error                      0.4333
Root mean squared error                  0.5252
Relative absolute error                 76.8147 %
Root relative squared error             80.7067 %
Total Number of Instances              585
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
```

```
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7583
Mean absolute error                      0.2421
Root mean squared error                  0.4268
Relative absolute error                 42.9172 %
Root relative squared error             65.5856 %
Total Number of Instances              585
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         386               65.9829 %
Incorrectly Classified Instances       199               34.0171 %
Kappa statistic                          0.4897
Mean absolute error                      0.316
Root mean squared error                  0.4098
Relative absolute error                 71.1088 %
Root relative squared error             86.9197 %
Coverage of cases (0.95 level)          91.7949 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.882    0.177    0.714      0.882    0.789      0.675
                  ↪  0.848     0.671     1
                 0.246    0.077    0.615      0.246    0.352      0.235
                  ↪  0.596     0.409     2
                 0.851    0.256    0.624      0.851    0.720      0.563
                  ↪  0.803     0.589     3
Weighted Avg.    0.660    0.170    0.651      0.660    0.620      0.491
 ↪  0.749     0.556

=== Confusion Matrix ===

   a   b   c   <-- classified as
 172  15   8 |   a = 1
  55  48  92 |   b = 2
  14  15 166 |   c = 3
```

```
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         497               84.9573 %
Incorrectly Classified Instances        88               15.0427 %
Kappa statistic                          0.7744
Mean absolute error                      0.1495
Root mean squared error                  0.2888
Relative absolute error                 33.6349 %
Root relative squared error             61.264  %
Coverage of cases (0.95 level)          97.6068 %
Mean rel. region size (0.95 level)      60.7407 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.923    0.056    0.891      0.923    0.907      0.859
                  ↪  0.944     0.886     1
                 0.810    0.113    0.782      0.810    0.796      0.691
                  ↪  0.885     0.802     2
                 0.815    0.056    0.878      0.815    0.846      0.774
                  ↪  0.940     0.846     3
Weighted Avg.    0.850    0.075    0.851      0.850    0.850      0.775
 ↪  0.923     0.844

=== Confusion Matrix ===

   a   b   c   <-- classified as
 180  10   5 |   a = 1
  20 158  17 |   b = 2
   2  34 159 |   c = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.6871
Mean absolute error                      0.4006
Root mean squared error                  0.5921
```

```
Relative absolute error                 65.6725 %
Root relative squared error             84.6233 %
Total Number of Instances              175
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         124               70.8571 %
Incorrectly Classified Instances        51               29.1429 %
Kappa statistic                          0.5625
Mean absolute error                      0.2313
Root mean squared error                  0.3587
Relative absolute error                 52.0384 %
Root relative squared error             76.0854 %
Coverage of cases (0.95 level)          96      %
Mean rel. region size (0.95 level)      68.9524 %
Total Number of Instances              175

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.845    0.103    0.803      0.845    0.824      0.733
                  ↪  0.923     0.861     1
                 0.579    0.153    0.647      0.579    0.611      0.440
                  ↪  0.805     0.642     2
                 0.700    0.183    0.667      0.700    0.683      0.512
                  ↪  0.875     0.806     3
Weighted Avg.    0.709    0.146    0.706      0.709    0.706      0.562
 ↪  0.868     0.771

=== Confusion Matrix ===

  a  b  c   <-- classified as
 49  3  6 |  a = 1
  9 33 15 |  b = 2
  3 15 42 |  c = 3
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt
```

```
Correctly Classified Instances         493               84.2735 %
Incorrectly Classified Instances        92               15.7265 %
Kappa statistic                          0.7641
Mean absolute error                      0.198
Root mean squared error                  0.2956
Relative absolute error                 44.5432 %
Root relative squared error             62.7033 %
Coverage of cases (0.95 level)          97.9487 %
Mean rel. region size (0.95 level)      70.4274 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.933    0.072    0.867      0.933    0.899      0.847
                  ↪  0.962     0.901     1
                 0.831    0.126    0.768      0.831    0.798      0.692
                  ↪  0.903     0.820     2
                 0.764    0.038    0.909      0.764    0.830      0.762
                  ↪  0.944     0.880     3
Weighted Avg.    0.843    0.079    0.848      0.843    0.842      0.767
 ↪  0.936     0.867

=== Confusion Matrix ===

   a   b   c   <-- classified as
 182  10   3 |   a = 1
  21 162  12 |   b = 2
   7  39 149 |   c = 3
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7596
Mean absolute error                      0.2726
Root mean squared error                  0.4232
Relative absolute error                 48.323  %
Root relative squared error             65.0352 %
Total Number of Instances              585
```

D.1.4 combined-exp-excl0-noresample-export

```
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         372               63.5897 %
Incorrectly Classified Instances       213               36.4103 %
Kappa statistic                          0.3708
Mean absolute error                      0.2879
Root mean squared error                  0.4053
Relative absolute error                 73.6275 %
Root relative squared error             91.6975 %
Coverage of cases (0.95 level)          96.2393 %
Mean rel. region size (0.95 level)      72.3647 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.743    0.313    0.606      0.743    0.668      0.421
                  ↪  0.793     0.636     1
                 0.598    0.306    0.659      0.598    0.627      0.293
                  ↪  0.712     0.704     2
                 0.422    0.023    0.692      0.422    0.524      0.499
                  ↪  0.872     0.521     3
Weighted Avg.    0.636    0.278    0.642      0.636    0.632      0.366
 ↪  0.761     0.657

=== Confusion Matrix ===

   a   b   c   <-- classified as
 171  57   2 |   a = 1
 107 174  10 |   b = 2
   4  33  27 |   c = 3
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         291               49.7436 %
Incorrectly Classified Instances       294               50.2564 %
Kappa statistic                          0
```

```
Mean absolute error                      0.391
Root mean squared error                  0.442
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.496     0.391     1
                 1.000    1.000    0.497      1.000    0.664      0.000
                  ↪  0.494     0.494     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.480     0.106     3
Weighted Avg.    0.497    0.497    0.247      0.497    0.330      0.000
 ↪  0.494     0.411

=== Confusion Matrix ===

   a   b   c   <-- classified as
   0 230   0 |   a = 1
   0 291   0 |   b = 2
   0  64   0 |   c = 3
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         486               83.0769 %
Incorrectly Classified Instances        99               16.9231 %
Kappa statistic                          0.7045
Mean absolute error                      0.1562
Root mean squared error                  0.3119
Relative absolute error                 39.9459 %
Root relative squared error             70.5797 %
Coverage of cases (0.95 level)          93.8462 %
Mean rel. region size (0.95 level)      55.5556 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.904    0.113    0.839      0.904    0.870      0.782
                  ↪  0.932     0.877     1
                 0.856    0.167    0.836      0.856    0.846      0.689
                  ↪  0.880     0.871     2
                 0.453    0.019    0.744      0.453    0.563      0.543
                  ↪  0.860     0.554     3
Weighted Avg.    0.831    0.129    0.827      0.831    0.824      0.710
 ↪  0.898     0.839

=== Confusion Matrix ===

   a   b   c   <-- classified as
 208  21   1 |   a = 1
  33 249   9 |   b = 2
   7  28  29 |   c = 3
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         393               67.1795 %
Incorrectly Classified Instances       192               32.8205 %
Kappa statistic                          0.4135
Mean absolute error                      0.3066
Root mean squared error                  0.398
Relative absolute error                 78.4052 %
Root relative squared error             90.0496 %
Coverage of cases (0.95 level)          94.8718 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.843    0.270    0.669      0.843    0.746      0.560
                  ↪  0.802     0.636     1
                 0.674    0.310    0.683      0.674    0.678      0.364
                  ↪  0.681     0.622     2
                 0.047    0.010    0.375      0.047    0.083      0.100
                  ↪  0.750     0.279     3
Weighted Avg.    0.672    0.261    0.644      0.672    0.640      0.412
 ↪  0.736     0.590
```

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 194  35   1 |   a = 1
  91 196   4 |   b = 2
   5  56   3 |   c = 3
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.1184
Mean absolute error                      0.5641
Root mean squared error                  0.6507
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              585
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         135               77.1429 %
Incorrectly Classified Instances        40               22.8571 %
Kappa statistic                          0.6153
Mean absolute error                      0.1959
Root mean squared error                  0.3621
Relative absolute error                 49.1596 %
Root relative squared error             79.4523 %
Coverage of cases (0.95 level)          90.2857 %
Mean rel. region size (0.95 level)      60.9524 %
Total Number of Instances              175

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.944    0.250    0.720      0.944    0.817      0.683
                  ↪  0.822     0.639     1
                 0.756    0.124    0.831      0.756    0.792      0.640
                  ↪  0.824     0.795     2
                 0.346    0.013    0.818      0.346    0.486      0.488
                  ↪  0.837     0.621     3
```

```
Weighted Avg.    0.771    0.159    0.784      0.771    0.757      0.635
 ↪  0.825     0.706

=== Confusion Matrix ===

  a  b  c   <-- classified as
 67  4  0 |  a = 1
 17 59  2 |  b = 2
  9  8  9 |  c = 3
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correlation coefficient                  0.5887
Mean absolute error                      0.4333
Root mean squared error                  0.5252
Relative absolute error                 76.8147 %
Root relative squared error             80.7067 %
Total Number of Instances              585
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7583
Mean absolute error                      0.2421
Root mean squared error                  0.4268
Relative absolute error                 42.9172 %
Root relative squared error             65.5856 %
Total Number of Instances              585
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt


Correctly Classified Instances         482               82.3932 %
Incorrectly Classified Instances       103               17.6068 %
Kappa statistic                          0.6943
```

```
Mean absolute error                      0.1766
Root mean squared error                  0.3138
Relative absolute error                 45.1691 %
Root relative squared error             70.9975 %
Coverage of cases (0.95 level)          96.7521 %
Mean rel. region size (0.95 level)      68.433  %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.891    0.110    0.840      0.891    0.865      0.774
                  ↪  0.924     0.874     1
                 0.852    0.163    0.838      0.852    0.845      0.689
                  ↪  0.867     0.851     2
                 0.453    0.031    0.644      0.453    0.532      0.495
                  ↪  0.849     0.457     3
Weighted Avg.    0.824    0.128    0.818      0.824    0.819      0.701
 ↪  0.887     0.817

=== Confusion Matrix ===

   a   b   c   <-- classified as
 205  20   5 |   a = 1
  32 248  11 |   b = 2
   7  28  29 |   c = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.6871
Mean absolute error                      0.4006
Root mean squared error                  0.5921
Relative absolute error                 65.6725 %
Root relative squared error             84.6233 %
Total Number of Instances              175
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filt
```

```
Correctly Classified Instances         483               82.5641 %
Incorrectly Classified Instances       102               17.4359 %
Kappa statistic                          0.6911
Mean absolute error                      0.1892
Root mean squared error                  0.3042
Relative absolute error                 48.3954 %
Root relative squared error             68.8317 %
Coverage of cases (0.95 level)          97.9487 %
Mean rel. region size (0.95 level)      70.9402 %
Total Number of Instances              585

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.891    0.096    0.858      0.891    0.874      0.790
                  ↪  0.936     0.897     1
                 0.880    0.207    0.808      0.880    0.842      0.675
                  ↪  0.893     0.868     2
                 0.344    0.013    0.759      0.344    0.473      0.475
                  ↪  0.874     0.526     3
Weighted Avg.    0.826    0.142    0.822      0.826    0.814      0.698
 ↪  0.908     0.842

=== Confusion Matrix ===

   a   b   c   <-- classified as
 205  25   0 |   a = 1
  28 256   7 |   b = 2
   6  36  22 |   c = 3
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7596
Mean absolute error                      0.2726
Root mean squared error                  0.4232
Relative absolute error                 48.323  %
Root relative squared error             65.0352 %
Total Number of Instances              585
```

D.1.5 combined-exp-excl0-linreg

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8
Relation:
 ↪  combined-exp-excl0-data-weka.filters.unsupervised.attribute.Remove-R3
Instances:    585
Attributes:   3
              numpeople
              numactive
              sizeconnected
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

numpeople =

      0.0456 * numactive +
     -0.024  * sizeconnected +
      1.1772

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.5887
Mean absolute error                      0.4333
Root mean squared error                  0.5252
Relative absolute error                 76.8147 %
Root relative squared error             80.7067 %
Total Number of Instances              585
```

D.1.6 subexp1-all-export

```
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         103               85.8333 %
Incorrectly Classified Instances        17               14.1667 %
Kappa statistic                          0.7167
Mean absolute error                      0.2281
Root mean squared error                  0.3417
Relative absolute error                 45.6123 %
Root relative squared error             68.3499 %
Coverage of cases (0.95 level)          99.1667 %
Mean rel. region size (0.95 level)      85      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.850    0.133    0.864      0.850    0.857      0.717
                  ↪  0.850     0.856     0
                 0.867    0.150    0.852      0.867    0.860      0.717
                  ↪  0.850     0.783     1
Weighted Avg.    0.858    0.142    0.858      0.858    0.858      0.717
 ↪  0.850     0.819

=== Confusion Matrix ===

  a  b   <-- classified as
 51  9 |  a = 0
  8 52 |  b = 1
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          60               50      %
Incorrectly Classified Instances        60               50      %
Kappa statistic                          0
Mean absolute error                      0.5
Root mean squared error                  0.5
```

```
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.500      1.000    0.667      0.000
                  ↪  0.500     0.500     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.500     1
Weighted Avg.    0.500    0.500    0.250      0.500    0.333      0.000
 ↪  0.500     0.500

=== Confusion Matrix ===

  a  b   <-- classified as
 60  0 |  a = 0
 60  0 |  b = 1
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          97               80.8333 %
Incorrectly Classified Instances        23               19.1667 %
Kappa statistic                          0.6167
Mean absolute error                      0.2003
Root mean squared error                  0.3914
Relative absolute error                 40.0572 %
Root relative squared error             78.274  %
Coverage of cases (0.95 level)          99.1667 %
Mean rel. region size (0.95 level)      67.0833 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.883    0.267    0.768      0.883    0.822      0.624
                  ↪  0.920     0.900     0
                 0.733    0.117    0.863      0.733    0.793      0.624
                  ↪  0.920     0.928     1
```

```
Weighted Avg.    0.808    0.192    0.815      0.808    0.807      0.624
 ↪  0.920     0.914

=== Confusion Matrix ===

  a  b   <-- classified as
 53  7 |  a = 0
 16 44 |  b = 1
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          68               56.6667 %
Incorrectly Classified Instances        52               43.3333 %
Kappa statistic                          0.1333
Mean absolute error                      0.4216
Root mean squared error                  0.5844
Relative absolute error                 84.3128 %
Root relative squared error            116.8817 %
Coverage of cases (0.95 level)          74.1667 %
Mean rel. region size (0.95 level)      67.9167 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.250    0.117    0.682      0.250    0.366      0.172
                  ↪  0.740     0.732     0
                 0.883    0.750    0.541      0.883    0.671      0.172
                  ↪  0.740     0.825     1
Weighted Avg.    0.567    0.433    0.611      0.567    0.518      0.172
 ↪  0.740     0.779

=== Confusion Matrix ===

  a  b   <-- classified as
 15 45 |  a = 0
  7 53 |  b = 1
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
```

```
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          96               80      %
Incorrectly Classified Instances        24               20      %
Kappa statistic                          0.6
Mean absolute error                      0.2
Root mean squared error                  0.4472
Relative absolute error                 40      %
Root relative squared error             89.4427 %
Coverage of cases (0.95 level)          80      %
Mean rel. region size (0.95 level)      50      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.400    0.714      1.000    0.833      0.655
                  ↪  0.800     0.714     0
                 0.600    0.000    1.000      0.600    0.750      0.655
                  ↪  0.800     0.800     1
Weighted Avg.    0.800    0.200    0.857      0.800    0.792      0.655
 ↪  0.800     0.757

=== Confusion Matrix ===

  a  b   <-- classified as
 60  0 |  a = 0
 24 36 |  b = 1
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          99               82.5    %
Incorrectly Classified Instances        21               17.5    %
Kappa statistic                          0.65
Mean absolute error                      0.1916
Root mean squared error                  0.305
Relative absolute error                 38.3222 %
Root relative squared error             60.9961 %
Coverage of cases (0.95 level)          99.1667 %
Mean rel. region size (0.95 level)      70      %
Total Number of Instances              120
```

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.767    0.117    0.868      0.767    0.814      0.654
                  ↪  0.923     0.897     0
                 0.883    0.233    0.791      0.883    0.835      0.654
                  ↪  0.923     0.929     1
Weighted Avg.    0.825    0.175    0.829      0.825    0.824      0.654
 ↪  0.923     0.913

=== Confusion Matrix ===

  a  b   <-- classified as
 46 14 |  a = 0
  7 53 |  b = 1
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.5986
Mean absolute error                      0.1914
Root mean squared error                  0.3792
Relative absolute error                 55.6068 %
Root relative squared error             86.7056 %
Total Number of Instances               36
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.2975
Mean absolute error                      0.331
Root mean squared error                  0.4091
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              121
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
```

```
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.7446
Mean absolute error                      0.1466
Root mean squared error                  0.2704
Relative absolute error                 44.2753 %
Root relative squared error             66.1036 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.0855
Mean absolute error                      0.5707
Root mean squared error                  0.7263
Relative absolute error                172.4024 %
Root relative squared error            177.5205 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7762
Mean absolute error                      0.1216
Root mean squared error                  0.2555
Relative absolute error                 36.7494 %
Root relative squared error             62.4413 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp1-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          28               77.7778 %
Incorrectly Classified Instances         8               22.2222 %
```

```
Kappa statistic                          0.52
Mean absolute error                      0.2133
Root mean squared error                  0.3453
Relative absolute error                 41.7968 %
Root relative squared error             67.3913 %
Coverage of cases (0.95 level)          97.2222 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances               36

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.864    0.357    0.792      0.864    0.826      0.524
                  ↪  0.890     0.902     0
                 0.643    0.136    0.750      0.643    0.692      0.524
                  ↪  0.890     0.823     1
Weighted Avg.    0.778    0.271    0.775      0.778    0.774      0.524
 ↪  0.890     0.871

=== Confusion Matrix ===

  a  b   <-- classified as
 19  3 |  a = 0
  5  9 |  b = 1
```

D.1.7 subexp2-all-export

```
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          82               68.3333 %
Incorrectly Classified Instances        38               31.6667 %
Kappa statistic                          0.525
Mean absolute error                      0.2314
Root mean squared error                  0.4216
Relative absolute error                 52.0737 %
Root relative squared error             89.4389 %
Coverage of cases (0.95 level)          82.5    %
Mean rel. region size (0.95 level)      50.2778 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.325    0.000    1.000      0.325    0.491      0.493
                  ↪  0.880     0.766     0
                 0.850    0.400    0.515      0.850    0.642      0.426
                  ↪  0.839     0.789     1
                 0.875    0.075    0.854      0.875    0.864      0.795
                  ↪  0.964     0.914     2
Weighted Avg.    0.683    0.158    0.790      0.683    0.665      0.572
 ↪  0.894     0.823

=== Confusion Matrix ===

  a  b  c   <-- classified as
 13 27  0 |  a = 0
  0 34  6 |  b = 1
  0  5 35 |  c = 2
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          40               33.3333 %
```

```
Incorrectly Classified Instances        80               66.6667 %
Kappa statistic                          0
Mean absolute error                      0.4444
Root mean squared error                  0.4714
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.333      1.000    0.500      0.000
                  ↪  0.500     0.333     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.333     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.333     2
Weighted Avg.    0.333    0.333    0.111      0.333    0.167      0.000
 ↪  0.500     0.333

=== Confusion Matrix ===

  a  b  c   <-- classified as
 40  0  0 |  a = 0
 40  0  0 |  b = 1
 40  0  0 |  c = 2
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          91               75.8333 %
Incorrectly Classified Instances        29               24.1667 %
Kappa statistic                          0.6375
Mean absolute error                      0.2504
Root mean squared error                  0.3594
Relative absolute error                 56.3474 %
Root relative squared error             76.2441 %
Coverage of cases (0.95 level)          90.8333 %
Mean rel. region size (0.95 level)      76.6667 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.850    0.025    0.944      0.850    0.895      0.849
                  ↪  0.827     0.801     0
                 0.650    0.188    0.634      0.650    0.642      0.460
                  ↪  0.803     0.603     1
                 0.775    0.150    0.721      0.775    0.747      0.614
                  ↪  0.907     0.869     2
Weighted Avg.    0.758    0.121    0.767      0.758    0.761      0.641
 ↪  0.846     0.758

=== Confusion Matrix ===

  a  b  c   <-- classified as
 34  6  0 |  a = 0
  2 26 12 |  b = 1
  0  9 31 |  c = 2
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         110               91.6667 %
Incorrectly Classified Instances        10                8.3333 %
Kappa statistic                          0.875
Mean absolute error                      0.1413
Root mean squared error                  0.2488
Relative absolute error                 31.7905 %
Root relative squared error             52.7684 %
Coverage of cases (0.95 level)          99.1667 %
Mean rel. region size (0.95 level)      58.6111 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.850    0.013    0.971      0.850    0.907      0.869
                  ↪  0.959     0.918     0
                 0.900    0.075    0.857      0.900    0.878      0.815
                  ↪  0.923     0.884     1
                 1.000    0.038    0.930      1.000    0.964      0.946
                  ↪  0.978     0.932     2
Weighted Avg.    0.917    0.042    0.920      0.917    0.916      0.877
 ↪  0.953     0.911
```

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 34  6  0 |  a = 0
  1 36  3 |  b = 1
  0  0 40 |  c = 2
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         104               86.6667 %
Incorrectly Classified Instances        16               13.3333 %
Kappa statistic                          0.8
Mean absolute error                      0.149
Root mean squared error                  0.2703
Relative absolute error                 33.5265 %
Root relative squared error             57.3481 %
Coverage of cases (0.95 level)          99.1667 %
Mean rel. region size (0.95 level)      61.6667 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.775    0.013    0.969      0.775    0.861      0.813
                  ↪  0.950     0.882     0
                 0.825    0.088    0.825      0.825    0.825      0.738
                  ↪  0.898     0.804     1
                 1.000    0.100    0.833      1.000    0.909      0.866
                  ↪  0.981     0.950     2
Weighted Avg.    0.867    0.067    0.876      0.867    0.865      0.805
 ↪  0.943     0.879

=== Confusion Matrix ===

  a  b  c   <-- classified as
 31  7  2 |  a = 0
  1 33  6 |  b = 1
  0  0 40 |  c = 2
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
```

```
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          69               57.5    %
Incorrectly Classified Instances        51               42.5    %
Kappa statistic                          0.3625
Mean absolute error                      0.3519
Root mean squared error                  0.4472
Relative absolute error                 79.1667 %
Root relative squared error             94.8683 %
Coverage of cases (0.95 level)          88.3333 %
Mean rel. region size (0.95 level)      68.3333 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.825    0.388    0.516      0.825    0.635      0.413
                  ↪  0.696     0.459     0
                 0.700    0.188    0.651      0.700    0.675      0.504
                  ↪  0.781     0.581     1
                 0.200    0.063    0.615      0.200    0.302      0.209
                  ↪  0.579     0.394     2
Weighted Avg.    0.575    0.213    0.594      0.575    0.537      0.375
 ↪  0.685     0.478

=== Confusion Matrix ===

  a  b  c   <-- classified as
 33  6  1 |  a = 0
  8 28  4 |  b = 1
 23  9  8 |  c = 2
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.5766
Mean absolute error                      0.5049
Root mean squared error                  0.7069
Relative absolute error                 73.9302 %
Root relative squared error             92.7733 %
Total Number of Instances              121
```

```
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.2601
Mean absolute error                      0.6829
Root mean squared error                  0.7619
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              121
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.5421
Mean absolute error                      0.4539
Root mean squared error                  0.6352
Relative absolute error                 66.4667 %
Root relative squared error             83.3663 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8302
Mean absolute error                      0.2803
Root mean squared error                  0.4224
Relative absolute error                 41.0495 %
Root relative squared error             55.4409 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst
```

```
Correlation coefficient                  0.8413
Mean absolute error                      0.3713
Root mean squared error                  0.4604
Relative absolute error                 59.6445 %
Root relative squared error             65.6294 %
Total Number of Instances               36
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp2-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          31               86.1111 %
Incorrectly Classified Instances         5               13.8889 %
Kappa statistic                          0.7917
Mean absolute error                      0.1461
Root mean squared error                  0.2745
Relative absolute error                 32.7407 %
Root relative squared error             57.9718 %
Coverage of cases (0.95 level)          94.4444 %
Mean rel. region size (0.95 level)      60.1852 %
Total Number of Instances               36

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.833    0.000    1.000      0.833    0.909      0.877
                  ↪  0.990     0.967     0
                 0.786    0.091    0.846      0.786    0.815      0.705
                  ↪  0.880     0.863     1
                 1.000    0.115    0.769      1.000    0.870      0.825
                  ↪  0.912     0.777     2
Weighted Avg.    0.861    0.067    0.876      0.861    0.861      0.796
 ↪  0.925     0.874

=== Confusion Matrix ===

  a  b  c   <-- classified as
 10  2  0 |  a = 0
  0 11  3 |  b = 1
  0  0 10 |  c = 2
```

D.1.8 subexp3-all-export

```
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances        102               85      %
Incorrectly Classified Instances       18               15      %
Kappa statistic                         0.8
Mean absolute error                     0.127
Root mean squared error                 0.2896
Relative absolute error                33.8677 %
Root relative squared error            66.8708 %
Coverage of cases (0.95 level)         89.1667 %
Mean rel. region size (0.95 level)     38.125  %
Total Number of Instances             120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.044    0.882      1.000    0.938      0.918
                  ↪  0.961     0.788     0
                 0.833    0.033    0.893      0.833    0.862      0.819
                  ↪  0.888     0.811     1
                 0.833    0.089    0.758      0.833    0.794      0.722
                  ↪  0.897     0.692     2
                 0.733    0.033    0.880      0.733    0.800      0.746
                  ↪  0.929     0.868     3
Weighted Avg.    0.850    0.050    0.853      0.850    0.848      0.801
 ↪  0.919     0.790

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 30  0  0  0 |  a = 0
  4 25  0  1 |  b = 1
  0  3 25  2 |  c = 2
  0  0  8 22 |  d = 3
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns
```

```
Correctly Classified Instances          98               81.6667 %
Incorrectly Classified Instances        22               18.3333 %
Kappa statistic                          0.7556
Mean absolute error                      0.1635
Root mean squared error                  0.278
Relative absolute error                 43.5893 %
Root relative squared error             64.2077 %
Coverage of cases (0.95 level)          95.8333 %
Mean rel. region size (0.95 level)      62.0833 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.800    0.044    0.857      0.800    0.828      0.774
                  ↪  0.961     0.776     0
                 0.800    0.056    0.828      0.800    0.814      0.753
                  ↪  0.928     0.892     1
                 0.867    0.122    0.703      0.867    0.776      0.698
                  ↪  0.888     0.675     2
                 0.800    0.022    0.923      0.800    0.857      0.817
                  ↪  0.910     0.779     3
Weighted Avg.    0.817    0.061    0.828      0.817    0.819      0.761
 ↪  0.922     0.781

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 24  3  3  0 |  a = 0
  4 24  2  0 |  b = 1
  0  2 26  2 |  c = 2
  0  0  6 24 |  d = 3
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          30               25      %
Incorrectly Classified Instances        90               75      %
Kappa statistic                          0
Mean absolute error                      0.375
Root mean squared error                  0.433
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
```

```
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.250      1.000    0.400      0.000
                  ↪  0.500     0.250     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     3
Weighted Avg.    0.250    0.250    0.063      0.250    0.100      0.000
 ↪  0.500     0.250

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 30  0  0  0 |  a = 0
 30  0  0  0 |  b = 1
 30  0  0  0 |  c = 2
 30  0  0  0 |  d = 3
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          76               63.3333 %
Incorrectly Classified Instances        44               36.6667 %
Kappa statistic                          0.5111
Mean absolute error                      0.1998
Root mean squared error                  0.3745
Relative absolute error                 53.2876 %
Root relative squared error             86.4948 %
Coverage of cases (0.95 level)          75      %
Mean rel. region size (0.95 level)      41.25   %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
```

```
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.981     0.879     0
                 0.867    0.044    0.867      0.867    0.867      0.822
                  ↪  0.958     0.932     1
                 0.867    0.144    0.667      0.867    0.754      0.668
                  ↪  0.915     0.729     2
                 0.800    0.300    0.471      0.800    0.593      0.438
                  ↪  0.879     0.847     3
Weighted Avg.    0.633    0.122    0.501      0.633    0.553      0.482
 ↪  0.933     0.847

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  0  2  6 22 |  a = 0
  0 26  1  3 |  b = 1
  0  2 26  2 |  c = 2
  0  0  6 24 |  d = 3
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances        105               87.5    %
Incorrectly Classified Instances       15               12.5    %
Kappa statistic                         0.8333
Mean absolute error                     0.1429
Root mean squared error                 0.2597
Relative absolute error                38.0977 %
Root relative squared error            59.9645 %
Coverage of cases (0.95 level)         98.3333 %
Mean rel. region size (0.95 level)     56.4583 %
Total Number of Instances             120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.044    0.882      1.000    0.938      0.918
                  ↪  0.968     0.844     0
                 0.833    0.022    0.926      0.833    0.877      0.841
                  ↪  0.945     0.925     1
                 0.867    0.078    0.788      0.867    0.825      0.765
                  ↪  0.918     0.727     2
                 0.800    0.022    0.923      0.800    0.857      0.817
                  ↪  0.932     0.885     3
```

```
Weighted Avg.    0.875    0.042    0.880      0.875    0.874      0.835
 ↪  0.941     0.845

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 30  0  0  0 |  a = 0
  4 25  1  0 |  b = 1
  0  2 26  2 |  c = 2
  0  0  6 24 |  d = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                0.8644
Mean absolute error                    0.3633
Root mean squared error                0.4996
Relative absolute error               42.2731 %
Root relative squared error           50.6432 %
Total Number of Instances             36
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                0.4037
Mean absolute error                    0.6513
Root mean squared error                1.0444
Relative absolute error               73.4525 %
Root relative squared error          103.4679 %
Total Number of Instances            121
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient               -0.193
Mean absolute error                    0.8866
Root mean squared error                1.0094
```

```
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              121
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.7851
Mean absolute error                      0.4287
Root mean squared error                  0.6226
Relative absolute error                 48.3559 %
Root relative squared error             61.6761 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8095
Mean absolute error                      0.405
Root mean squared error                  0.5975
Relative absolute error                 45.6761 %
Root relative squared error             59.1961 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          85               70.8333 %
Incorrectly Classified Instances        35               29.1667 %
Kappa statistic                          0.6111
Mean absolute error                      0.2965
Root mean squared error                  0.3785
Relative absolute error                 79.0741 %
Root relative squared error             87.4184 %
Coverage of cases (0.95 level)          90      %
Mean rel. region size (0.95 level)      75.4167 %
```

```
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.533    0.067    0.727      0.533    0.615      0.522
                  ↪  0.624     0.501     0
                 0.767    0.200    0.561      0.767    0.648      0.517
                  ↪  0.823     0.527     1
                 0.733    0.100    0.710      0.733    0.721      0.627
                  ↪  0.855     0.606     2
                 0.800    0.022    0.923      0.800    0.857      0.817
                  ↪  0.929     0.834     3
Weighted Avg.    0.708    0.097    0.730      0.708    0.710      0.621
 ↪  0.808     0.617

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 16 11  3  0 |  a = 0
  6 23  1  0 |  b = 1
  0  6 22  2 |  c = 2
  0  1  5 24 |  d = 3
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp3-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          30               83.3333 %
Incorrectly Classified Instances         6               16.6667 %
Kappa statistic                          0.7662
Mean absolute error                      0.1368
Root mean squared error                  0.2721
Relative absolute error                 35.7485 %
Root relative squared error             61.3624 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      54.1667 %
Total Number of Instances               36

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.857    0.069    0.750      0.857    0.800      0.750
                  ↪  0.951     0.706     0
```

```
                 0.867    0.095    0.867      0.867    0.867      0.771
                  ↪  0.940     0.902     1
                 0.857    0.069    0.750      0.857    0.800      0.750
                  ↪  0.936     0.679     2
                 0.714    0.000    1.000      0.714    0.833      0.817
                  ↪  0.961     0.868     3
Weighted Avg.    0.833    0.067    0.847      0.833    0.834      0.772
 ↪  0.945     0.814

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  6  1  0  0 |  a = 0
  2 13  0  0 |  b = 1
  0  1  6  0 |  c = 2
  0  0  2  5 |  d = 3
```

D.1.9 subexp4-all-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          21               25.9259 %
Incorrectly Classified Instances        60               74.0741 %
Kappa statistic                         -0.1111
Mean absolute error                      0.4451
Root mean squared error                  0.4721
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.333    0.444    0.273      0.333    0.300      -0.107
                  ↪  0.438     0.308     0
                 0.222    0.333    0.250      0.222    0.235      -0.115
                  ↪  0.438     0.308     1
                 0.222    0.333    0.250      0.222    0.235      -0.115
                  ↪  0.438     0.308     2
Weighted Avg.    0.259    0.370    0.258      0.259    0.257      -0.112
 ↪  0.438     0.308

=== Confusion Matrix ===

  a  b  c   <-- classified as
  9  9  9 |  a = 0
 12  6  9 |  b = 1
 12  9  6 |  c = 2
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          47               58.0247 %
Incorrectly Classified Instances        34               41.9753 %
```

```
Kappa statistic                          0.3704
Mean absolute error                      0.3423
Root mean squared error                  0.4121
Relative absolute error                 76.903  %
Root relative squared error             87.2952 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      90.535  %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.074    0.000    1.000      0.074    0.138      0.225
                  ↪  0.505     0.316     0
                 1.000    0.574    0.466      1.000    0.635      0.445
                  ↪  0.694     0.441     1
                 0.667    0.056    0.857      0.667    0.750      0.657
                  ↪  0.730     0.767     2
Weighted Avg.    0.580    0.210    0.774      0.580    0.508      0.443
 ↪  0.643     0.508

=== Confusion Matrix ===

  a  b  c   <-- classified as
  2 22  3 |  a = 0
  0 27  0 |  b = 1
  0  9 18 |  c = 2
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          47               58.0247 %
Incorrectly Classified Instances        34               41.9753 %
Kappa statistic                          0.3704
Mean absolute error                      0.3025
Root mean squared error                  0.4351
Relative absolute error                 67.9599 %
Root relative squared error             92.1633 %
Coverage of cases (0.95 level)          86.4198 %
Mean rel. region size (0.95 level)      60.9053 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.074    0.000    1.000      0.074    0.138      0.225
                  ↪  0.664     0.565     0
                 1.000    0.574    0.466      1.000    0.635      0.445
                  ↪  0.715     0.458     1
                 0.667    0.056    0.857      0.667    0.750      0.657
                  ↪  0.726     0.750     2
Weighted Avg.    0.580    0.210    0.774      0.580    0.508      0.443
 ↪  0.702     0.591

=== Confusion Matrix ===

  a  b  c   <-- classified as
  2 22  3 |  a = 0
  0 27  0 |  b = 1
  0  9 18 |  c = 2
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          45               55.5556 %
Incorrectly Classified Instances        36               44.4444 %
Kappa statistic                          0.3333
Mean absolute error                      0.3381
Root mean squared error                  0.4249
Relative absolute error                 75.9641 %
Root relative squared error             90.0007 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      83.5391 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.492     0.312     0
                 1.000    0.519    0.491      1.000    0.659      0.486
                  ↪  0.693     0.442     1
                 0.667    0.148    0.692      0.667    0.679      0.524
                  ↪  0.722     0.750     2
Weighted Avg.    0.556    0.222    0.394      0.556    0.446      0.337
 ↪  0.636     0.501
```

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
  0 19  8 |  a = 0
  0 27  0 |  b = 1
  0  9 18 |  c = 2
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          47               58.0247 %
Incorrectly Classified Instances        34               41.9753 %
Kappa statistic                          0.3704
Mean absolute error                      0.3403
Root mean squared error                  0.4136
Relative absolute error                 76.4565 %
Root relative squared error             87.6059 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      91.7695 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.074    0.000    1.000      0.074    0.138      0.225
                  ↪  0.507     0.317     0
                 1.000    0.574    0.466      1.000    0.635      0.445
                  ↪  0.693     0.441     1
                 0.667    0.056    0.857      0.667    0.750      0.657
                  ↪  0.733     0.769     2
Weighted Avg.    0.580    0.210    0.774      0.580    0.508      0.443
 ↪  0.644     0.509

=== Confusion Matrix ===

  a  b  c   <-- classified as
  2 22  3 |  a = 0
  0 27  0 |  b = 1
  0  9 18 |  c = 2
=== Evaluation result ===

Scheme: SMO Nom : SMO
```

```
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          45               55.5556 %
Incorrectly Classified Instances        36               44.4444 %
Kappa statistic                          0.3333
Mean absolute error                      0.3457
Root mean squared error                  0.4444
Relative absolute error                 77.6642 %
Root relative squared error             94.1361 %
Coverage of cases (0.95 level)          88.8889 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.333     0
                 1.000    0.630    0.443      1.000    0.614      0.405
                  ↪  0.685     0.443     1
                 0.667    0.037    0.900      0.667    0.766      0.688
                  ↪  0.815     0.711     2
Weighted Avg.    0.556    0.222    0.448      0.556    0.460      0.364
 ↪  0.667     0.496

=== Confusion Matrix ===

  a  b  c   <-- classified as
  0 25  2 |  a = 0
  0 27  0 |  b = 1
  0  9 18 |  c = 2
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.4834
Mean absolute error                      0.8114
Root mean squared error                  0.8684
Relative absolute error                 91.2026 %
Root relative squared error             92.0524 %
```

```
Total Number of Instances               24
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          14               58.3333 %
Incorrectly Classified Instances        10               41.6667 %
Kappa statistic                          0.3909
Mean absolute error                      0.3482
Root mean squared error                  0.3991
Relative absolute error                 78.1795 %
Root relative squared error             84.4528 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      90.2778 %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.588    0.412      1.000    0.583      0.412
                  ↪  0.580     0.340     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.813     0.571     1
                 0.778    0.000    1.000      0.778    0.875      0.828
                  ↪  0.867     0.861     2
Weighted Avg.    0.583    0.172    0.495      0.583    0.498      0.431
 ↪  0.765     0.612

=== Confusion Matrix ===

 a b c   <-- classified as
 7 0 0 | a = 0
 8 0 0 | b = 1
 2 0 7 | c = 2
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.4147
```

```
Mean absolute error                      0.7156
Root mean squared error                  0.9342
Relative absolute error                 74.9779 %
Root relative squared error             94.7967 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.5625
Mean absolute error                      0.6777
Root mean squared error                  0.8059
Relative absolute error                 71.0093 %
Root relative squared error             81.7794 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.4005
Mean absolute error                      0.9544
Root mean squared error                  0.9855
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               81
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp4-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.5296
Mean absolute error                      0.6956
Root mean squared error                  0.8209
Relative absolute error                 72.8793 %
Root relative squared error             83.299  %
Total Number of Instances               81
```

D.1.10 subexp5-all-export

```
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          40               50      %
Incorrectly Classified Instances        40               50      %
Kappa statistic                          0
Mean absolute error                      0.4836
Root mean squared error                  0.6512
Relative absolute error                 96.7263 %
Root relative squared error            130.2472 %
Coverage of cases (0.95 level)          65      %
Mean rel. region size (0.95 level)      62.5    %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.200    0.200    0.500      0.200    0.286      0.000
                  ↪  0.760     0.648     0
                 0.800    0.800    0.500      0.800    0.615      0.000
                  ↪  0.760     0.836     1
Weighted Avg.    0.500    0.500    0.500      0.500    0.451      0.000
 ↪  0.760     0.742

=== Confusion Matrix ===

  a  b   <-- classified as
  8 32 |  a = 0
  8 32 |  b = 1
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          64               80      %
Incorrectly Classified Instances        16               20      %
Kappa statistic                          0.6
Mean absolute error                      0.3283
```

```
Root mean squared error                  0.4732
Relative absolute error                 65.6533 %
Root relative squared error             94.631  %
Coverage of cases (0.95 level)          92.5    %
Mean rel. region size (0.95 level)      90      %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.825    0.225    0.786      0.825    0.805      0.601
                  ↪  0.697     0.688     0
                 0.775    0.175    0.816      0.775    0.795      0.601
                  ↪  0.697     0.645     1
Weighted Avg.    0.800    0.200    0.801      0.800    0.800      0.601
 ↪  0.697     0.667

=== Confusion Matrix ===

  a  b   <-- classified as
 33  7 |  a = 0
  9 31 |  b = 1
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          40               50      %
Incorrectly Classified Instances        40               50      %
Kappa statistic                          0
Mean absolute error                      0.5
Root mean squared error                  0.5
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.500      1.000    0.667      0.000
                  ↪  0.500     0.500     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.500     1
```

```
Weighted Avg.    0.500    0.500    0.250      0.500    0.333      0.000
 ↪  0.500     0.500

=== Confusion Matrix ===

  a  b   <-- classified as
 40  0 |  a = 0
 40  0 |  b = 1
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          67               83.75   %
Incorrectly Classified Instances        13               16.25   %
Kappa statistic                          0.675
Mean absolute error                      0.3301
Root mean squared error                  0.407
Relative absolute error                 66.0195 %
Root relative squared error             81.3976 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      89.375  %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.325    0.755      1.000    0.860      0.714
                  ↪  0.733     0.672     0
                 0.675    0.000    1.000      0.675    0.806      0.714
                  ↪  0.733     0.841     1
Weighted Avg.    0.838    0.163    0.877      0.838    0.833      0.714
 ↪  0.733     0.757

=== Confusion Matrix ===

  a  b   <-- classified as
 40  0 |  a = 0
 13 27 |  b = 1
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
```

```
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          51               63.75   %
Incorrectly Classified Instances        29               36.25   %
Kappa statistic                          0.275
Mean absolute error                      0.3625
Root mean squared error                  0.6021
Relative absolute error                 72.5    %
Root relative squared error            120.4159 %
Coverage of cases (0.95 level)          63.75   %
Mean rel. region size (0.95 level)      50      %
Total Number of Instances               80

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.475    0.200    0.704      0.475    0.567      0.291
                  ↪  0.638     0.597     0
                 0.800    0.525    0.604      0.800    0.688      0.291
                  ↪  0.638     0.583     1
Weighted Avg.    0.638    0.363    0.654      0.638    0.628      0.291
 ↪  0.638     0.590

=== Confusion Matrix ===

  a  b   <-- classified as
 19 21 |  a = 0
  8 32 |  b = 1
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          51               63.75   %
Incorrectly Classified Instances        29               36.25   %
Kappa statistic                          0.275
Mean absolute error                      0.3976
Root mean squared error                  0.456
Relative absolute error                 79.5265 %
Root relative squared error             91.2064 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances               80
```

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.650    0.375    0.634      0.650    0.642      0.275
                  ↪  0.675     0.562     0
                 0.625    0.350    0.641      0.625    0.633      0.275
                  ↪  0.675     0.797     1
Weighted Avg.    0.638    0.363    0.638      0.638    0.637      0.275
 ↪  0.675     0.680

=== Confusion Matrix ===

  a  b   <-- classified as
 26 14 |  a = 0
 15 25 |  b = 1
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.2199
Mean absolute error                      0.4128
Root mean squared error                  0.5929
Relative absolute error                 81.6719 %
Root relative squared error            116.9857 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.6618
Mean absolute error                      0.2767
Root mean squared error                  0.3766
Relative absolute error                 54.7362 %
Root relative squared error             74.3137 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
```

```
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.3637
Mean absolute error                      0.5054
Root mean squared error                  0.5068
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               81
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.6101
Mean absolute error                      0.3683
Root mean squared error                  0.3986
Relative absolute error                 73.992  %
Root relative squared error             79.9633 %
Total Number of Instances               24
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.5335
Mean absolute error                      0.3532
Root mean squared error                  0.4223
Relative absolute error                 69.8874 %
Root relative squared error             83.3216 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp5-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          19               79.1667 %
Incorrectly Classified Instances         5               20.8333 %
Kappa statistic                          0.5652
```

```
Mean absolute error                      0.4139
Root mean squared error                  0.4663
Relative absolute error                 82.5464 %
Root relative squared error             92.9393 %
Coverage of cases (0.95 level)          95.8333 %
Mean rel. region size (0.95 level)      97.9167 %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.455    0.722      1.000    0.839      0.628
                  ↪  0.608     0.633     0
                 0.545    0.000    1.000      0.545    0.706      0.628
                  ↪  0.608     0.738     1
Weighted Avg.    0.792    0.246    0.850      0.792    0.778      0.628
 ↪  0.608     0.681

=== Confusion Matrix ===

  a  b   <-- classified as
 13  0 |  a = 0
  5  6 |  b = 1
```

D.1.11 subexp6-all-export

```
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          94               78.3333 %
Incorrectly Classified Instances        26               21.6667 %
Kappa statistic                          0.675
Mean absolute error                      0.1651
Root mean squared error                  0.3523
Relative absolute error                 37.1455 %
Root relative squared error             74.7332 %
Coverage of cases (0.95 level)          83.3333 %
Mean rel. region size (0.95 level)      46.3889 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.500    0.013    0.952      0.500    0.656      0.605
                  ↪  0.903     0.857     0
                 0.875    0.263    0.625      0.875    0.729      0.579
                  ↪  0.904     0.811     1
                 0.975    0.050    0.907      0.975    0.940      0.909
                  ↪  0.965     0.946     2
Weighted Avg.    0.783    0.108    0.828      0.783    0.775      0.698
 ↪  0.924     0.871

=== Confusion Matrix ===

  a  b  c   <-- classified as
 20 20  0 |  a = 0
  1 35  4 |  b = 1
  0  1 39 |  c = 2
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         109               90.8333 %
```

```
Incorrectly Classified Instances        11                9.1667 %
Kappa statistic                          0.8625
Mean absolute error                      0.1352
Root mean squared error                  0.2756
Relative absolute error                 30.4194 %
Root relative squared error             58.454  %
Coverage of cases (0.95 level)          94.1667 %
Mean rel. region size (0.95 level)      54.4444 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.925    0.000    1.000      0.925    0.961      0.944
                  ↪  0.945     0.956     0
                 0.825    0.013    0.971      0.825    0.892      0.850
                  ↪  0.924     0.903     1
                 0.975    0.125    0.796      0.975    0.876      0.815
                  ↪  0.933     0.805     2
Weighted Avg.    0.908    0.046    0.922      0.908    0.910      0.870
 ↪  0.934     0.888

=== Confusion Matrix ===

  a  b  c   <-- classified as
 37  0  3 |  a = 0
  0 33  7 |  b = 1
  0  1 39 |  c = 2
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          40               33.3333 %
Incorrectly Classified Instances        80               66.6667 %
Kappa statistic                          0
Mean absolute error                      0.4444
Root mean squared error                  0.4714
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              120

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.333      1.000    0.500      0.000
                  ↪  0.500     0.333     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.333     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.333     2
Weighted Avg.    0.333    0.333    0.111      0.333    0.167      0.000
 ↪  0.500     0.333

=== Confusion Matrix ===

  a  b  c   <-- classified as
 40  0  0 |  a = 0
 40  0  0 |  b = 1
 40  0  0 |  c = 2
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         110               91.6667 %
Incorrectly Classified Instances        10                8.3333 %
Kappa statistic                          0.875
Mean absolute error                      0.1268
Root mean squared error                  0.2454
Relative absolute error                 28.5193 %
Root relative squared error             52.0613 %
Coverage of cases (0.95 level)          97.5    %
Mean rel. region size (0.95 level)      58.8889 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.925    0.013    0.974      0.925    0.949      0.925
                  ↪  0.991     0.981     0
                 0.875    0.063    0.875      0.875    0.875      0.813
                  ↪  0.941     0.931     1
                 0.950    0.050    0.905      0.950    0.927      0.889
                  ↪  0.962     0.894     2
Weighted Avg.    0.917    0.042    0.918      0.917    0.917      0.876
 ↪  0.965     0.935
```

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 37  3  0 |  a = 0
  1 35  4 |  b = 1
  0  2 38 |  c = 2
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances        102               85      %
Incorrectly Classified Instances       18               15      %
Kappa statistic                         0.775
Mean absolute error                     0.162
Root mean squared error                 0.2669
Relative absolute error                36.4527 %
Root relative squared error            56.6234 %
Coverage of cases (0.95 level)         99.1667 %
Mean rel. region size (0.95 level)     74.7222 %
Total Number of Instances             120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.700    0.013    0.966      0.700    0.812      0.757
                  ↪  0.965     0.953     0
                 0.875    0.138    0.761      0.875    0.814      0.715
                  ↪  0.904     0.803     1
                 0.975    0.075    0.867      0.975    0.918      0.876
                  ↪  0.941     0.837     2
Weighted Avg.    0.850    0.075    0.864      0.850    0.848      0.783
 ↪  0.937     0.864

=== Confusion Matrix ===

  a  b  c   <-- classified as
 28 10  2 |  a = 0
  1 35  4 |  b = 1
  0  1 39 |  c = 2
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
```

```
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          97               80.8333 %
Incorrectly Classified Instances        23               19.1667 %
Kappa statistic                          0.7125
Mean absolute error                      0.2907
Root mean squared error                  0.3776
Relative absolute error                 65.4167 %
Root relative squared error             80.1041 %
Coverage of cases (0.95 level)          88.3333 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances              120

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.600    0.000    1.000      0.600    0.750      0.707
                  ↪  0.788     0.748     0
                 0.850    0.175    0.708      0.850    0.773      0.650
                  ↪  0.838     0.652     1
                 0.975    0.113    0.813      0.975    0.886      0.830
                  ↪  0.938     0.805     2
Weighted Avg.    0.808    0.096    0.840      0.808    0.803      0.729
 ↪  0.854     0.735

=== Confusion Matrix ===

  a  b  c   <-- classified as
 24 13  3 |  a = 0
  0 34  6 |  b = 1
  0  1 39 |  c = 2
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.7812
Mean absolute error                      0.2955
Root mean squared error                  0.531
Relative absolute error                 42.4203 %
Root relative squared error             64.0541 %
Total Number of Instances              121
```

```
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.2731
Mean absolute error                      0.6966
Root mean squared error                  0.8289
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              121
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.7865
Mean absolute error                      0.3703
Root mean squared error                  0.5076
Relative absolute error                 53.1549 %
Root relative squared error             61.2347 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8562
Mean absolute error                      0.2513
Root mean squared error                  0.4318
Relative absolute error                 36.0772 %
Root relative squared error             52.0928 %
Total Number of Instances              121
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns
```

```
Correctly Classified Instances          34               94.4444 %
Incorrectly Classified Instances         2                5.5556 %
Kappa statistic                          0.9157
Mean absolute error                      0.1518
Root mean squared error                  0.2324
Relative absolute error                 34.0162 %
Root relative squared error             49.066  %
Coverage of cases (0.95 level)          97.2222 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances               36

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    0.042    0.923      1.000    0.960      0.941
                  ↪  0.986     0.974     0
                 0.929    0.045    0.929      0.929    0.929      0.883
                  ↪  0.977     0.971     1
                 0.900    0.000    1.000      0.900    0.947      0.931
                  ↪  0.973     0.959     2
Weighted Avg.    0.944    0.032    0.947      0.944    0.944      0.916
 ↪  0.979     0.969

=== Confusion Matrix ===

  a  b  c   <-- classified as
 12  0  0 |  a = 0
  1 13  0 |  b = 1
  0  1  9 |  c = 2
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp6-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.9413
Mean absolute error                      0.1944
Root mean squared error                  0.3058
Relative absolute error                 27.0353 %
Root relative squared error             36.3823 %
Total Number of Instances               36
```

D.1.12 subexp7-all-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          30               23.4375 %
Incorrectly Classified Instances        98               76.5625 %
Kappa statistic                         -0.0208
Mean absolute error                      0.3752
Root mean squared error                  0.4333
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.750    0.813    0.235      0.750    0.358      -0.067
                  ↪  0.463     0.236     0
                 0.188    0.208    0.231      0.188    0.207      -0.022
                  ↪  0.463     0.236     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.463     0.236     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.463     0.236     3
Weighted Avg.    0.234    0.255    0.117      0.234    0.141      -0.022
 ↪  0.463     0.236

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 24  8  0  0 |  a = 0
 26  6  0  0 |  b = 1
 26  6  0  0 |  c = 2
 26  6  0  0 |  d = 3
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns
```

```
Correctly Classified Instances          71               55.4688 %
Incorrectly Classified Instances        57               44.5313 %
Kappa statistic                          0.4063
Mean absolute error                      0.2453
Root mean squared error                  0.3955
Relative absolute error                 65.3732 %
Root relative squared error             91.2797 %
Coverage of cases (0.95 level)          79.6875 %
Mean rel. region size (0.95 level)      46.6797 %
Total Number of Instances              128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.798     0.438     0
                 0.750    0.146    0.632      0.750    0.686      0.573
                  ↪  0.875     0.585     1
                 0.750    0.302    0.453      0.750    0.565      0.394
                  ↪  0.840     0.603     2
                 0.719    0.146    0.622      0.719    0.667      0.547
                  ↪  0.906     0.765     3
Weighted Avg.    0.555    0.148    0.427      0.555    0.479      0.378
 ↪  0.855     0.598

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
  0 11 13  8 |  a = 0
  0 24  7  1 |  b = 1
  0  3 24  5 |  c = 2
  0  0  9 23 |  d = 3
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          81               63.2813 %
Incorrectly Classified Instances        47               36.7188 %
Kappa statistic                          0.5104
Mean absolute error                      0.2165
Root mean squared error                  0.3814
Relative absolute error                 57.6909 %
Root relative squared error             88.0312 %
```

```
Coverage of cases (0.95 level)          96.0938 %
Mean rel. region size (0.95 level)      49.0234 %
Total Number of Instances              128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.906    0.208    0.592      0.906    0.716      0.622
                  ↪  0.896     0.676     0
                 0.188    0.063    0.500      0.188    0.273      0.186
                  ↪  0.774     0.517     1
                 0.875    0.198    0.596      0.875    0.709      0.608
                  ↪  0.868     0.547     2
                 0.563    0.021    0.900      0.563    0.692      0.646
                  ↪  0.911     0.829     3
Weighted Avg.    0.633    0.122    0.647      0.633    0.597      0.515
 ↪  0.862     0.642

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 29  3  0  0 |  a = 0
 20  6  5  1 |  b = 1
  0  3 28  1 |  c = 2
  0  0 14 18 |  d = 3
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          94               73.4375 %
Incorrectly Classified Instances        34               26.5625 %
Kappa statistic                          0.6458
Mean absolute error                      0.1818
Root mean squared error                  0.318
Relative absolute error                 48.4522 %
Root relative squared error             73.404  %
Coverage of cases (0.95 level)          96.0938 %
Mean rel. region size (0.95 level)      59.7656 %
Total Number of Instances              128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
```

```
                 0.625    0.021    0.909      0.625    0.741      0.693
                  ↪  0.903     0.732     0
                 0.813    0.135    0.667      0.813    0.732      0.637
                  ↪  0.849     0.620     1
                 0.781    0.167    0.610      0.781    0.685      0.570
                  ↪  0.881     0.635     2
                 0.719    0.031    0.885      0.719    0.793      0.740
                  ↪  0.927     0.857     3
Weighted Avg.    0.734    0.089    0.768      0.734    0.738      0.660
 ↪  0.890     0.711

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 20  8  4  0 |  a = 0
  2 26  3  1 |  b = 1
  0  5 25  2 |  c = 2
  0  0  9 23 |  d = 3
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances        100               78.125  %
Incorrectly Classified Instances       28               21.875  %
Kappa statistic                         0.7083
Mean absolute error                     0.1771
Root mean squared error                 0.299
Relative absolute error                47.2065 %
Root relative squared error            69.0101 %
Coverage of cases (0.95 level)         98.4375 %
Mean rel. region size (0.95 level)     48.6328 %
Total Number of Instances             128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.656    0.021    0.913      0.656    0.764      0.717
                  ↪  0.896     0.701     0
                 0.781    0.125    0.676      0.781    0.725      0.627
                  ↪  0.907     0.725     1
                 0.906    0.125    0.707      0.906    0.795      0.725
                  ↪  0.927     0.784     2
                 0.781    0.021    0.926      0.781    0.847      0.807
                  ↪  0.957     0.898     3
```

```
Weighted Avg.    0.781    0.073    0.805      0.781    0.783      0.719
 ↪  0.922     0.777

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 21 11  0  0 |  a = 0
  2 25  5  0 |  b = 1
  0  1 29  2 |  c = 2
  0  0  7 25 |  d = 3
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.2087
Mean absolute error                      0.8936
Root mean squared error                  1.0137
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              130
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.8117
Mean absolute error                      0.4321
Root mean squared error                  0.5891
Relative absolute error                 48.3546 %
Root relative squared error             58.1189 %
Total Number of Instances              130
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.5921
Mean absolute error                      0.6652
Root mean squared error                  0.8986
```

```
Relative absolute error                 74.4371 %
Root relative squared error             88.6532 %
Total Number of Instances              130
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8173
Mean absolute error                      0.4017
Root mean squared error                  0.586
Relative absolute error                 44.9546 %
Root relative squared error             57.813  %
Total Number of Instances              130
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          82               64.0625 %
Incorrectly Classified Instances        46               35.9375 %
Kappa statistic                          0.5208
Mean absolute error                      0.2982
Root mean squared error                  0.3803
Relative absolute error                 79.4721 %
Root relative squared error             87.7815 %
Coverage of cases (0.95 level)          93.75   %
Mean rel. region size (0.95 level)      76.1719 %
Total Number of Instances              128

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.531    0.031    0.850      0.531    0.654      0.596
                  ↪  0.781    0.639     0
                 0.563    0.073    0.720      0.563    0.632      0.535
                  ↪  0.726    0.528     1
                 0.969    0.375    0.463      0.969    0.626      0.515
                  ↪  0.804    0.460     2
                 0.500    0.000    1.000      0.500    0.667      0.655
                  ↪  0.927    0.767     3
```

```
Weighted Avg.    0.641    0.120    0.758      0.641    0.645      0.575
 ↪  0.809     0.598

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 17  7  8  0 |  a = 0
  2 18 12  0 |  b = 1
  1  0 31  0 |  c = 2
  0  0 16 16 |  d = 3
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          22               57.8947 %
Incorrectly Classified Instances        16               42.1053 %
Kappa statistic                          0.4432
Mean absolute error                      0.2648
Root mean squared error                  0.3885
Relative absolute error                 70.5409 %
Root relative squared error             89.6012 %
Coverage of cases (0.95 level)          97.3684 %
Mean rel. region size (0.95 level)      71.0526 %
Total Number of Instances               38

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.556    0.034    0.833      0.556    0.667      0.608
                  ↪  0.847     0.766     0
                 0.667    0.103    0.667      0.667    0.667      0.563
                  ↪  0.743     0.567     1
                 0.667    0.414    0.333      0.667    0.444      0.215
                  ↪  0.678     0.389     2
                 0.455    0.000    1.000      0.455    0.625      0.610
                  ↪  0.963     0.903     3
Weighted Avg.    0.579    0.131    0.724      0.579    0.602      0.505
 ↪  0.816     0.669

=== Confusion Matrix ===

 a b c d   <-- classified as
 5 0 4 0 | a = 0
 1 6 2 0 | b = 1
```

```
 0 3 6 0 | c = 2
 0 0 6 5 | d = 3
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp7-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.69
Mean absolute error                      0.6367
Root mean squared error                  0.7897
Relative absolute error                 73.8176 %
Root relative squared error             80.4426 %
Total Number of Instances               39
```

D.1.13 subexp8-all-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          40               25      %
Incorrectly Classified Instances       120               75      %
Kappa statistic                          0
Mean absolute error                      0.375
Root mean squared error                  0.433
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 1.000    1.000    0.250      1.000    0.400      0.000
                  ↪  0.500     0.250     0
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     1
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     2
                 0.000    0.000    0.000      0.000    0.000      0.000
                  ↪  0.500     0.250     3
Weighted Avg.    0.250    0.250    0.063      0.250    0.100      0.000
 ↪  0.500     0.250

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 40  0  0  0 |  a = 0
 40  0  0  0 |  b = 1
 40  0  0  0 |  c = 2
 40  0  0  0 |  d = 3
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns
```

```
Correctly Classified Instances         123               76.875  %
Incorrectly Classified Instances        37               23.125  %
Kappa statistic                          0.6917
Mean absolute error                      0.1799
Root mean squared error                  0.3064
Relative absolute error                 47.9674 %
Root relative squared error             70.756  %
Coverage of cases (0.95 level)          97.5    %
Mean rel. region size (0.95 level)      52.5    %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.900    0.083    0.783      0.900    0.837      0.781
                  ↪  0.925     0.798     0
                 0.750    0.083    0.750      0.750    0.750      0.667
                  ↪  0.948     0.863     1
                 0.750    0.092    0.732      0.750    0.741      0.653
                  ↪  0.917     0.822     2
                 0.675    0.050    0.818      0.675    0.740      0.669
                  ↪  0.936     0.769     3
Weighted Avg.    0.769    0.077    0.771      0.769    0.767      0.692
 ↪  0.931     0.813

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 36  4  0  0 |  a = 0
 10 30  0  0 |  b = 1
  0  4 30  6 |  c = 2
  0  2 11 27 |  d = 3
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         111               69.375  %
Incorrectly Classified Instances        49               30.625  %
Kappa statistic                          0.5917
Mean absolute error                      0.1637
Root mean squared error                  0.3375
Relative absolute error                 43.6485 %
Root relative squared error             77.9332 %
```

```
Coverage of cases (0.95 level)          83.125  %
Mean rel. region size (0.95 level)      37.0313 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.300    0.000    1.000      0.300    0.462      0.493
                  ↪  0.942     0.832     0
                 0.925    0.075    0.804      0.925    0.860      0.813
                  ↪  0.980     0.933     1
                 0.650    0.183    0.542      0.650    0.591      0.441
                  ↪  0.870     0.789     2
                 0.900    0.150    0.667      0.900    0.766      0.687
                  ↪  0.924     0.807     3
Weighted Avg.    0.694    0.102    0.753      0.694    0.670      0.609
 ↪  0.929     0.840

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 12  9 16  3 |  a = 0
  0 37  2  1 |  b = 1
  0  0 26 14 |  c = 2
  0  0  4 36 |  d = 3
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         124               77.5    %
Incorrectly Classified Instances        36               22.5    %
Kappa statistic                          0.7
Mean absolute error                      0.1866
Root mean squared error                  0.2876
Relative absolute error                 49.7652 %
Root relative squared error             66.4218 %
Coverage of cases (0.95 level)          98.125  %
Mean rel. region size (0.95 level)      64.8438 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
```

```
                 0.400    0.000    1.000      0.400    0.571      0.577
                  ↪  0.835     0.660     0
                 1.000    0.158    0.678      1.000    0.808      0.755
                  ↪  0.963     0.817     1
                 0.825    0.075    0.786      0.825    0.805      0.738
                  ↪  0.923     0.836     2
                 0.875    0.067    0.814      0.875    0.843      0.790
                  ↪  0.930     0.759     3
Weighted Avg.    0.775    0.075    0.819      0.775    0.757      0.715
 ↪  0.913     0.768

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 16 19  4  1 |  a = 0
  0 40  0  0 |  b = 1
  0  0 33  7 |  c = 2
  0  0  5 35 |  d = 3
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         141               88.125  %
Incorrectly Classified Instances        19               11.875  %
Kappa statistic                          0.8417
Mean absolute error                      0.1531
Root mean squared error                  0.2647
Relative absolute error                 40.8136 %
Root relative squared error             61.1334 %
Coverage of cases (0.95 level)          98.125  %
Mean rel. region size (0.95 level)      56.4063 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.750    0.000    1.000      0.750    0.857      0.832
                  ↪  0.936     0.812     0
                 1.000    0.083    0.800      1.000    0.889      0.856
                  ↪  0.982     0.947     1
                 0.825    0.017    0.943      0.825    0.880      0.847
                  ↪  0.953     0.847     2
                 0.950    0.058    0.844      0.950    0.894      0.859
                  ↪  0.955     0.852     3
```

```
Weighted Avg.    0.881    0.040    0.897      0.881    0.880      0.848
 ↪  0.957     0.865

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 30 10  0  0 |  a = 0
  0 40  0  0 |  b = 1
  0  0 33  7 |  c = 2
  0  0  2 38 |  d = 3
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          90               56.25   %
Incorrectly Classified Instances        70               43.75   %
Kappa statistic                          0.4167
Mean absolute error                      0.2958
Root mean squared error                  0.377
Relative absolute error                 78.8889 %
Root relative squared error             87.0558 %
Coverage of cases (0.95 level)          98.125  %
Mean rel. region size (0.95 level)      75.1563 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.300    0.133    0.429      0.300    0.353      0.190
                  ↪  0.715     0.430     0
                 0.600    0.083    0.706      0.600    0.649      0.547
                  ↪  0.893     0.626     1
                 0.900    0.333    0.474      0.900    0.621      0.491
                  ↪  0.805     0.463     2
                 0.450    0.033    0.818      0.450    0.581      0.524
                  ↪  0.891     0.675     3
Weighted Avg.    0.563    0.146    0.607      0.563    0.551      0.438
 ↪  0.826     0.549

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 12 10 18  0 |  a = 0
```

```
 16 24  0  0 |  b = 1
  0  0 36  4 |  c = 2
  0  0 22 18 |  d = 3
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.2047
Mean absolute error                      0.9163
Root mean squared error                  1.0428
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              161
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8516
Mean absolute error                      0.3818
Root mean squared error                  0.5853
Relative absolute error                 41.6675 %
Root relative squared error             56.1301 %
Total Number of Instances              161
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.7318
Mean absolute error                      0.5248
Root mean squared error                  0.7079
Relative absolute error                 57.2766 %
Root relative squared error             67.8882 %
Total Number of Instances              161
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
```

```
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8056
Mean absolute error                      0.4348
Root mean squared error                  0.6267
Relative absolute error                 47.455  %
Root relative squared error             60.0947 %
Total Number of Instances              161
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8945
Mean absolute error                      0.333
Root mean squared error                  0.4896
Relative absolute error                 35.8476 %
Root relative squared error             44.9663 %
Total Number of Instances               48
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp8-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          41                 85.4167 %
Incorrectly Classified Instances         7                 14.5833 %
Kappa statistic                          0.8053
Mean absolute error                      0.1055
Root mean squared error                  0.2462
Relative absolute error                 28.0495 %
Root relative squared error             56.6297 %
Coverage of cases (0.95 level)          93.75   %
Mean rel. region size (0.95 level)      43.2292 %
Total Number of Instances               48

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.800    0.000    1.000      0.800    0.889      0.856
                  ↪  0.992     0.980     0
```

```
                 1.000    0.081    0.786      1.000    0.880      0.850
                  ↪  0.990     0.963     1
                 0.833    0.056    0.833      0.833    0.833      0.778
                  ↪  0.904     0.788     2
                 0.800    0.053    0.800      0.800    0.800      0.747
                  ↪  0.938     0.729     3
Weighted Avg.    0.854    0.043    0.868      0.854    0.854      0.812
 ↪  0.958     0.876

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 12  3  0  0 |  a = 0
  0 11  0  0 |  b = 1
  0  0 10  2 |  c = 2
  0  0  2  8 |  d = 3
```

D.1.14 subexp9-all-export

```
=== Evaluation result ===

Scheme: ZeroR Nom : ZeroR
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          21               25.9259 %
Incorrectly Classified Instances        60               74.0741 %
Kappa statistic                         -0.1111
Mean absolute error                      0.4451
Root mean squared error                  0.4721
Relative absolute error                100      %
Root relative squared error            100      %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)     100      %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.333    0.444    0.273      0.333    0.300      -0.107
                  ↪  0.438     0.308     0
                 0.222    0.333    0.250      0.222    0.235      -0.115
                  ↪  0.438     0.308     1
                 0.222    0.333    0.250      0.222    0.235      -0.115
                  ↪  0.438     0.308     2
Weighted Avg.    0.259    0.370    0.258      0.259    0.257      -0.112
 ↪  0.438     0.308

=== Confusion Matrix ===

  a  b  c   <-- classified as
  9  9  9 |  a = 0
 12  6  9 |  b = 1
 12  9  6 |  c = 2
=== Evaluation result ===

Scheme: NaiveBayes Nom : NaiveBayes
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          73               90.1235 %
Incorrectly Classified Instances         8                9.8765 %
Kappa statistic                          0.8519
```

```
Mean absolute error                      0.0943
Root mean squared error                  0.2432
Relative absolute error                 21.1964 %
Root relative squared error             51.5105 %
Coverage of cases (0.95 level)          95.0617 %
Mean rel. region size (0.95 level)      41.9753 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.778    0.000    1.000      0.778    0.875      0.837
                  ↪  0.912     0.915     0
                 1.000    0.074    0.871      1.000    0.931      0.898
                  ↪  0.958     0.826     1
                 0.926    0.074    0.862      0.926    0.893      0.838
                  ↪  0.951     0.884     2
Weighted Avg.    0.901    0.049    0.911      0.901    0.900      0.857
 ↪  0.940     0.875

=== Confusion Matrix ===

  a  b  c   <-- classified as
 21  2  4 |  a = 0
  0 27  0 |  b = 1
  0  2 25 |  c = 2
=== Evaluation result ===

Scheme: J48 Nom : J48
Options: -C 0.25 -M 2
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          70               86.4198 %
Incorrectly Classified Instances        11               13.5802 %
Kappa statistic                          0.7963
Mean absolute error                      0.1126
Root mean squared error                  0.2504
Relative absolute error                 25.3059 %
Root relative squared error             53.0374 %
Coverage of cases (0.95 level)          97.5309 %
Mean rel. region size (0.95 level)      57.6132 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===
```

```
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.778    0.056    0.875      0.778    0.824      0.746
                  ↪  0.933     0.899     0
                 1.000    0.074    0.871      1.000    0.931      0.898
                  ↪  0.955     0.823     1
                 0.815    0.074    0.846      0.815    0.830      0.748
                  ↪  0.937     0.918     2
Weighted Avg.    0.864    0.068    0.864      0.864    0.862      0.797
 ↪  0.942     0.880

=== Confusion Matrix ===

  a  b  c   <-- classified as
 21  2  4 |  a = 0
  0 27  0 |  b = 1
  3  2 22 |  c = 2
=== Evaluation result ===

Scheme: KStar Nom : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          75               92.5926 %
Incorrectly Classified Instances         6                7.4074 %
Kappa statistic                          0.8889
Mean absolute error                      0.0939
Root mean squared error                  0.1956
Relative absolute error                 21.0888 %
Root relative squared error             41.4338 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      50.2058 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.852    0.000    1.000      0.852    0.920      0.891
                  ↪  0.992     0.988     0
                 1.000    0.074    0.871      1.000    0.931      0.898
                  ↪  0.951     0.800     1
                 0.926    0.037    0.926      0.926    0.926      0.889
                  ↪  0.977     0.970     2
Weighted Avg.    0.926    0.037    0.932      0.926    0.926      0.892
 ↪  0.973     0.919
```

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 23  2  2 |  a = 0
  0 27  0 |  b = 1
  0  2 25 |  c = 2
=== Evaluation result ===

Scheme: IBk Nom : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances         72               88.8889 %
Incorrectly Classified Instances        9               11.1111 %
Kappa statistic                         0.8333
Mean absolute error                     0.089
Root mean squared error                 0.2443
Relative absolute error                19.9918 %
Root relative squared error            51.7544 %
Coverage of cases (0.95 level)         95.0617 %
Mean rel. region size (0.95 level)     46.9136 %
Total Number of Instances              81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.778    0.000    1.000      0.778    0.875      0.837
                  ↪  0.968     0.954     0
                 1.000    0.130    0.794      1.000    0.885      0.831
                  ↪  0.945     0.806     1
                 0.889    0.037    0.923      0.889    0.906      0.860
                  ↪  0.976     0.967     2
Weighted Avg.    0.889    0.056    0.906      0.889    0.889      0.843
 ↪  0.963     0.909

=== Confusion Matrix ===

  a  b  c   <-- classified as
 21  4  2 |  a = 0
  0 27  0 |  b = 1
  0  3 24 |  c = 2
=== Evaluation result ===

Scheme: MLP Nom : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
```

```
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correctly Classified Instances          21               87.5    %
Incorrectly Classified Instances         3               12.5    %
Kappa statistic                          0.81
Mean absolute error                      0.1255
Root mean squared error                  0.2415
Relative absolute error                 28.1768 %
Root relative squared error             51.1033 %
Coverage of cases (0.95 level)         100      %
Mean rel. region size (0.95 level)      62.5    %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.714    0.000    1.000      0.714    0.833      0.799
                  ↪  1.000     1.000     0
                 1.000    0.063    0.889      1.000    0.941      0.913
                  ↪  0.957     0.854     1
                 0.889    0.133    0.800      0.889    0.842      0.742
                  ↪  0.967     0.951     2
Weighted Avg.    0.875    0.071    0.888      0.875    0.873      0.816
 ↪  0.973     0.933

=== Confusion Matrix ===

 a b c   <-- classified as
 5 0 2 | a = 0
 0 8 0 | b = 1
 0 1 8 | c = 2
=== Evaluation result ===

Scheme: IBk Num : IBk
Options: -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A
 ↪  \"weka.core.EuclideanDistance -R first-last\""
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.6925
Mean absolute error                      0.3881
Root mean squared error                  0.8285
Relative absolute error                 38.8752 %
Root relative squared error             81.9035 %
Total Number of Instances               81
```

```
=== Evaluation result ===

Scheme: KStar Num : KStar
Options: -B 20 -M a
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.8878
Mean absolute error                      0.2587
Root mean squared error                  0.472
Relative absolute error                 25.9105 %
Root relative squared error             46.6551 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: ZeroR Num : ZeroR
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                 -0.4578
Mean absolute error                      0.9984
Root mean squared error                  1.0116
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               81
=== Evaluation result ===

Scheme: LinearRegression Num : LinearRegression
Options: -S 0 -R 1.0E-8
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns


Correlation coefficient                  0.8164
Mean absolute error                      0.3699
Root mean squared error                  0.5702
Relative absolute error                 37.0529 %
Root relative squared error             56.3686 %
Total Number of Instances               81
=== Evaluation result ===

Scheme: SMO Nom : SMO
Options: -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K
 ↪  "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst-weka.filters.uns
```

```
Correctly Classified Instances          71               87.6543 %
Incorrectly Classified Instances        10               12.3457 %
Kappa statistic                          0.8148
Mean absolute error                      0.2606
Root mean squared error                  0.3354
Relative absolute error                 58.5564 %
Root relative squared error             71.0366 %
Coverage of cases (0.95 level)          95.0617 %
Mean rel. region size (0.95 level)      66.6667 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC
                  ↪  Area  PRC Area  Class
                 0.778    0.000    1.000      0.778    0.875      0.837
                  ↪  0.899     0.874     0
                 1.000    0.148    0.771      1.000    0.871      0.811
                  ↪  0.926     0.771     1
                 0.852    0.037    0.920      0.852    0.885      0.832
                  ↪  0.957     0.873     2
Weighted Avg.    0.877    0.062    0.897      0.877    0.877      0.826
 ↪  0.927     0.840

=== Confusion Matrix ===

  a  b  c   <-- classified as
 21  4  2 |  a = 0
  0 27  0 |  b = 1
  0  4 23 |  c = 2
=== Evaluation result ===

Scheme: MLP Num : MultilayerPerceptron
Options: -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
Relation:
 ↪  subexp9-all-data-weka.filters.unsupervised.attribute.ClassAssigner-Cfirst


Correlation coefficient                  0.9151
Mean absolute error                      0.2078
Root mean squared error                  0.406
Relative absolute error                 20.7769 %
Root relative squared error             40.5436 %
Total Number of Instances               24
```

APPENDIX E

Physical Form

To enable the prototype to be easily mounted on the ceiling, the prototype was placed on a flat board with feet that would enable it to be screwed into a pole, and the pole extended to jam the sensor against the ceiling and the floor using the pole (Figure E.2 on the next page, Figure E.1). Due to a wireless module and battery pack being added to the Raspberry Pi, it was feasible for the sensor to operate entirely wirelessly for several hours. However, in most cases it was more convenient to operate using wired power and Ethernet.

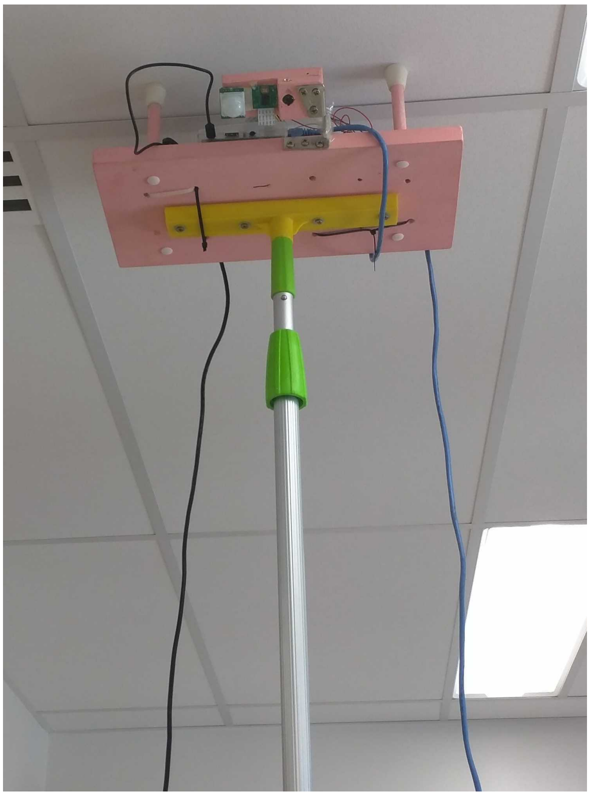
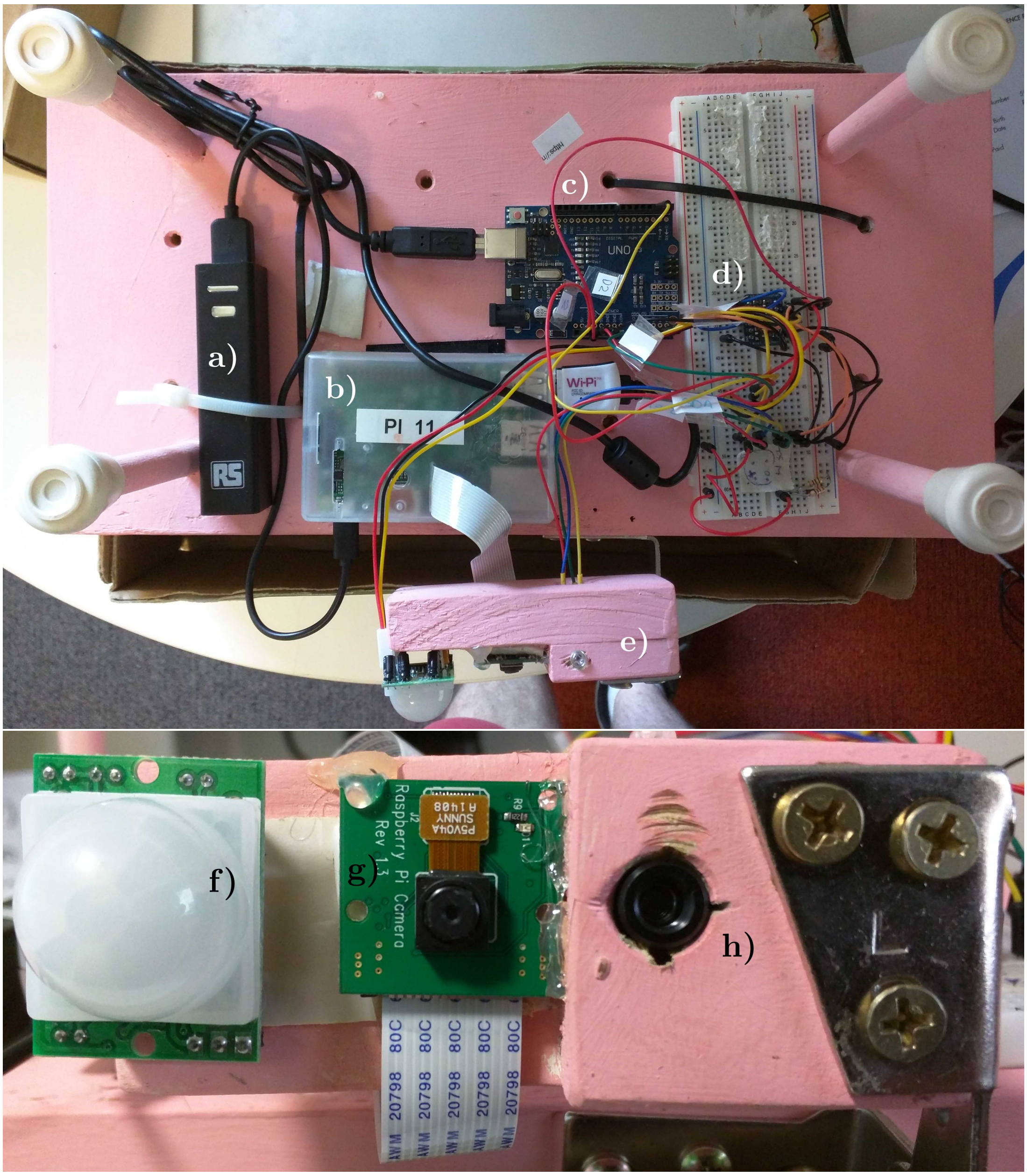


Figure E.1: Prototype in action



a) Battery pack

b) Raspberry Pi

c) Arduino

d) Level-shifting circuitry

e) Movable sensor mount

f) PIR

g) Camera

h) Melexis MLX90620 (*Melexis*)

Figure E.2: Prototype Physical Form

Bibliography

[1] Adafruit. 4-channel I2C-safe bi-directional logic level converter - BSS138 (product ID 757). `http://www.adafruit.com/product/757`. Accessed: 2015-01-07.

[2] Adafruit. PIR (motion) sensor (product ID 189). `http://www.adafruit.com/product/189`. Accessed: 2015-02-08.

[3] Arduino Forums. Arduino and MLX90620 16X4 pixel IR thermal array. `http://forum.arduino.cc/index.php/topic,126244.0.html`, 2012. Accessed: 2015-01-07.

[4] Atzori, L., Iera, A., and Morabito, G. The internet of things: A survey. *Computer networks 54*, 15 (2010), 2787–2805.

[5] Australian Bureau of Statistics. Household water and energy use, Victoria: Heating and cooling. Tech. Rep. 4602.2, October 2011. Retrieved October 6, 2014 from `http://www.abs.gov.au/ausstats/abs@.nsf/0/85424ADCCF6E5AE9CA257A670013AF89`.

[6] Balaji, B., Xu, J., Nwokafor, A., Gupta, R., and Agarwal, Y. Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (2013), ACM, p. 17.

[7] Beltran, A., Erickson, V. L., and Cerpa, A. E. ThermoSense: Occupancy thermal based sensing for HVAC control. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–8.

[8] Chan, M., Campo, E., Estève, D., and Fourniols, J.-Y. Smart homes - current features and future perspectives. *Maturitas 64*, 2 (2009), 90–97.

[9] Erickson, V. L., Achleitner, S., and Cerpa, A. E. POEM: Power-efficient occupancy-based energy management system. In *Proceedings of the 12th international conference on Information processing in sensor networks* (2013), ACM, pp. 203–216.

[10] Fisk, W. J., Faulkner, D., and Sullivan, D. P. Accuracy of CO2 sensors in commercial buildings: a pilot study. Tech. Rep. LBNL-61862, Lawrence Berkeley National Laboratory, 2006. Retrieved October 6, 2014 from `http://eaei.lbl.gov/sites/all/files/LBNL-61862_0.pdf`.

[11] Guinard, D., Ion, I., and Mayer, S. In search of an internet of things service architecture: REST or WS-*? a developers perspective. In *Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 2012, pp. 326–337.

[12] Guinard, D., Trifa, V., Mattern, F., and Wilde, E. From the internet of things to the web of things: Resource-oriented architecture and best practices. In *Architecting the Internet of Things*. Springer, 2011, pp. 97–129.

[13] Gupta, M., Intille, S. S., and Larson, K. Adding gps-control to traditional thermostats: An exploration of potential energy savings and design challenges. In *Pervasive Computing*. Springer, 2009, pp. 95–114.

[14] Hailemariam, E., Goldstein, R., Attar, R., and Khan, A. Real-time occupancy detection using decision trees with multiple sensor types. In *Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design* (2011), Society for Computer Simulation International, pp. 141–148.

[15] Hnat, T. W., Griffiths, E., Dawson, R., and Whitehouse, K. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems* (2012), ACM, pp. 309–322.

[16] Kleiminger, W., Beckel, C., Dey, A., and Santini, S. Inferring household occupancy patterns from unlabelled sensor data. Tech. Rep. 795, ETH Zurich, 2013. Retrieved October 6, 2014 from `http://eaei.lbl.gov/sites/all/files/LBNL-61862_0.pdf`.

[17] Kleiminger, W., Beckel, C., Staake, T., and Santini, S. Occupancy detection from electricity consumption data. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–8.

[18] Kovatsch, M. CoAP for the web of things: from tiny resource-constrained devices to the web browser. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (2013), ACM, pp. 1495–1504.

[19] Li, N., Calis, G., and Becerik-Gerber, B. Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. *Automation in construction 24* (2012), 89–99.

[20] Melexis. Datasheet IR thermometer 16X4 sensor array MLX90620. http://www.melexis.com/Asset/Datasheet-IR-thermometer-16X4-sensor-array-MLX90620-DownloadLink-6099.aspx, 2012. Accessed: 2015-01-07.

[21] Palattella, M. R., Accettura, N., Vilajosana, X., Watteyne, T., Grieco, L. A., Boggia, G., and Dohler, M. Standardized protocol stack for the internet of (important) things. *Communications Surveys & Tutorials, IEEE 15*, 3 (2013), 1389–1406.

[22] Serrano-Cuerda, J., Castillo, J. C., Sokolova, M. V., and Fernández-Caballero, A. Efficient people counting from indoor overhead video camera. In *Trends in Practical Applications of Agents and Multiagent Systems*. Springer, 2013, pp. 129–137.

[23] Shelby, Z., and Bormann, C. *6LoWPAN: The wireless embedded Internet*, vol. 43. John Wiley & Sons, 2011.

[24] SparkFun. FLiR Dev Kit. https://www.sparkfun.com/products/13233. Accessed: 2015-04-08.

[25] Teixeira, T., Dublon, G., and Savvides, A. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity. Tech. rep., Embedded Networks and Applications Lab (ENALAB), Yale University, 2010. Retrieved October 6, 2014 from http://www.eng.yale.edu/enalab/publications/human_sensing_enalabWIP.pdf.

[26] University of Waikato. Weka. http://www.cs.waikato.ac.nz/ml/weka/. Accessed: 2015-03-10.

[27] Winter, T., Thubert, P., Cisco Systems, Brandt, A., et al. RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks. RFC 6550, Internet Engineering Task Force, March 2012. Retrieved October 6, 2014 from http://tools.ietf.org/html/rfc6550.