Awareness and Monitoring of Personal Environment to improve Quality of Living at Home for the Elderly

Nathan Jones

A thesis presented for the degree of
Master of Philosophy

School of Computer Science and Informatics
Cardiff University
August 2022

1 Summary

As the average age of the population in the UK gets higher caution needs to be taken to ensure that elderly and vulnerable people are not left behind and can access the care and support that they need. This study aims to assess whether basic household sensors paired with a machine learning system can be used to act as an early warning system for carers, reducing carer workload while ensuring an equal level of care for their patients. To achieve this we have designed an anomaly detection system using a threshold system that aims to detect unusual behaviour in the property.

Partnering with the Safehouse company we have access to two properties that have been outfitted with a range of environmental sensors. For the purpose of this study focusing on the temperature, humidity, light and motion in these property.

Three time series prediction methods have been compared, these are LSTM, Arima and Autoencoder. A mixture of several important metrics were used to measure performance, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R2, as well as expert feedback from the team at Safehouse.

While we found the LSTM had the highest accuracy when making predictions on the data, particularly when using a combined sensor approach, due to the potential time taken to process data regularly it is not suited for a real world system. For this reason we suggest the use of an Autoencoder to make predictions on the data, with extra parameters to assist in identifying important anomalies.

Further research involving direct contact with carers is needed to discover whether this system is sufficient to their needs. Stricter monitoring of the people within the properties is also necessary in order to decide whether the anomalies detected identify all relevant real world problems.

Contents

1 Summary **i**

2 Acknowledgments **xii**

3 Introduction **1**

4 Methodology **5**
4.1 Background 5
4.1.1 Machine Learning 5
4.2 Machine Learning Models 6
4.2.1 LSTM 6
4.2.2 AutoEncoder 6
4.2.3 ARIMA 7
4.3 Measuring Model Accuracy 7
4.3.1 MAE 8
4.3.2 RMSE 8
4.3.3 R2 8
4.3.4 Expert Feedback 9
4.4 Model Comparisons 9
4.4.1 LSTM 10
4.4.2 AutoEncoder 10
4.4.3 ARIMA 11
4.5 Anomaly Detection 11
4.6 User Interface 11

5 Related Works **13**
5.1 Similar Papers 13
5.2 Living Alone 14
5.3 Time Series Forecasting 14
5.4 Model Accuracy 16
5.5 Anomaly Detection 17
5.6 Sensors 17

6 Experiment Setup **19**
6.1 Description 19
6.2 Partner Company 19
6.3 Setup 19

6.4 Environmental Sensors 20
6.4.1 Safehouse Hub 20
6.4.2 Safehouse USB 21
6.4.3 PIR Sensor 22
6.5 Data Reporting 24
6.6 Participants 25
6.7 Property Map 25

7 Model Comparison 28
7.1 Model Consideration 28
7.2 Data 29
7.3 Configurations 36
7.3.1 LSTM 36
7.3.2 LSTM AutoEncoder 38
7.4 Comparison of models 39
7.4.1 Individual Sensor Results 39
7.4.2 Time Taken 41
7.5 Conclusion 41

8 Anomaly Detection With Single and Multiple Sensors 43
8.1 Detecting Anomalies 43
8.2 Thresholds 43
8.2.1 Threshold Conclusion 50
8.3 Single Sensor Anomaly Detection 50
8.3.1 Details 50
8.3.2 Temperature 51
8.3.3 Humidity 52
8.3.4 Light 53
8.3.5 Motion 54
8.4 Combined Anomaly Detection 55
8.4.1 Method 55
8.4.2 Temperature with Humidity 55
8.4.3 Temperature with Light 60
8.4.4 Temperature with Sound 61
8.4.5 Humidity with Temperature 62
8.4.6 Humidity with Light 64
8.4.7 Humidity with Sound 65
8.5 Conclusion 66

9 Anomaly Detection With Additional Parameters **67**
9.1 Additional Conditions . . . 67
9.1.1 Traffic Light Range . . . 67
9.1.2 Constant . . . 70
9.1.3 Difference in Values . . . 70
9.1.4 Percentage Change . . . 71
9.1.5 Outside Comparison . . . 71
9.1.6 Motion in the Dark . . . 71
9.2 Temperature Results . . . 72
9.2.1 Standard anomalies . . . 73
9.2.2 Amber and red anomalies . . . 73
9.2.3 Difference anomalies . . . 73
9.2.4 Constant anomalies . . . 74
9.2.5 Percentage anomalies . . . 74
9.2.6 Overall . . . 74
9.2.7 Specialised compared to standard . . . 76
9.3 Humidity Results . . . 77
9.3.1 Amber and red anomalies . . . 78
9.3.2 Outside anomalies . . . 78
9.3.3 Constant anomalies . . . 78
9.3.4 Percentage anomalies . . . 78
9.3.5 Overall . . . 78
9.3.6 Specialised compared to standard . . . 81
9.4 Light Results . . . 83
9.4.1 Overall . . . 84
9.5 Motion Results . . . 85
9.5.1 Overall . . . 87
9.6 Conclusion . . . 88

10 User Interface and Visualisation **89**
10.1 Design . . . 89
10.1.1 Front End . . . 89
10.1.2 Charts . . . 90
10.2 Pages . . . 92
10.3 Customisation . . . 94
10.3.1 Chart Customisation . . . 97
10.4 Dashboard UI . . . 99
10.4.1 Prototype Design . . . 99

10.4.2 Final Design . . . 101
10.4.3 Usage . . . 103
10.4.4 Privacy . . . 104
10.5 Conclusion . . . 104

11 Discussion **105**
11.1 Anomaly Detection . . . 105
11.2 Extra Conditions . . . 106
11.3 Measuring Effectiveness . . . 107
11.4 User Interface . . . 108
11.5 User Comfort . . . 109
11.6 Property Data . . . 110

12 Conclusion **112**

13 Bibliography **115**

14 Appendix **120**
14.1 Extra tests . . . 120
14.2 Data . . . 120
14.3 Results . . . 123

List of Figures

1 Safehouse hub . . . 20
2 Safehouse USB . . . 21
3 Safehouse PIR . . . 22
4 Floor Plan Key . . . 25
5 Property 1 Ground Floor plan . . . 26
6 Property 1 First Floor plan . . . 26
7 Property 2 Ground Floor plan . . . 27
8 Property 1 Bluetooth data . . . 29
9 Property 2 Bluetooth data . . . 30
10 Anomalies detected in Property 1 Bluetooth temperature data using threshold with high of 1.2 and low of 1 . . . 44
11 Anomalies detected in Property 1 Bluetooth temperature data using threshold with high of 1 and low of 0.8 . . . 44
12 Anomalies detected in Property 1 humidity data using threshold with high of 1.2 and low of 1 . . . 45
13 Anomalies detected in Property 1 humidity data using threshold with high of 1 and low of 0.8 . . . 45
14 Anomalies detected in Property 1 temperature data using threshold with high of 5% and low of 10% . . . 47
15 Anomalies detected in Property 1 temperature data using threshold with high of 15% and low of 25% . . . 48
16 Anomalies detected in Property 1 humidity data using threshold with high of 5% and low of 10% . . . 48
17 Anomalies detected in Property 1 humidity data using threshold with high of 15% and low of 25% . . . 49
18 Regular anomaly detection in property 1 temperature . . . 51
19 Regular anomaly detection in property 2 temperature . . . 51
20 Regular anomaly detection in property 1 humidity . . . 52
21 Regular anomaly detection in property 2 humidity . . . 52
22 Regular anomaly detection in property 1 light . . . 53
23 Regular anomaly detection in property 2 light . . . 53
24 Regular anomaly detection in property 1 motion . . . 54
25 Regular anomaly detection in property 2 motion . . . 54
26 Anomaly detection in property 1 temperature supplemented by humidity . . . 57

27 Anomaly detection in property 2 temperature supplemented by humidity 57
28 Traffic light range for temperature 68
29 Traffic light range for humidity 69
30 Traffic light range for light 69
31 Property 1 Temperature with all conditions 72
32 Property 2 Temperature with all conditions 72
33 Property 1 Humidity with all conditions 77
34 Property 2 Humidity with all conditions 77
35 Property 1 Light with all conditions 83
36 Property 2 Light with all conditions 83
37 Property 1 Motion with all conditions 85
38 Property 2 Motion with all conditions 85
39 Basic chart 91
40 Navigation panel 92
41 Basic page customisation 94
42 Detail page multiple charts 95
43 Multiple page customisation 96
44 Chart scale 97
45 Chart highlighting 98
46 Date buttons 98
47 Dashboard Prototype 99
48 Dashboard Final 101
49 Dashboard card green 101
50 Dashboard card amber 102
51 Property 1 Bluetooth data 121
52 Property 2 Bluetooth data 122
53 Regular anomaly detection in property 1 temperature 124
54 Regular anomaly detection in property 2 temperature 125
55 Regular anomaly detection in property 1 humidity 126
56 Regular anomaly detection in property 2 humidity 127
57 Regular anomaly detection in property 1 light 128
58 Regular anomaly detection in property 2 light 129
59 Regular anomaly detection in property 1 motion 130
60 Regular anomaly detection in property 2 motion 131
61 Anomaly detection in property 1 temperature supplemented by humidity 132

62 Anomaly detection in property 2 temperature supplemented by humidity 133
63 Property 1 Temperature with all conditions 134
64 Property 2 Temperature with all conditions 135
65 Property 1 Humidity with all conditions 136
66 Property 2 Humidity with all conditions 137
67 Property 1 Light with all conditions 138
68 Property 2 Light with all conditions 139
69 Property 1 Motion with all conditions 140
70 Property 2 Motion with all conditions 141

List of Tables

1 Table of the data tracked by the Safehouse USB Sensor 21
2 Table of the data tracked by the Safehouse PIR Sensor 23
3 Property 1 Living Room Temperature Test Data 32
4 Property 1 Living Room Humidity Test Data 33
5 Property 1 Living Room Light Test Data 33
6 Property 1 Kitchen Motion Test Data 33
7 Property 2 Living Room Temperature Test Data 34
8 Property 2 Living Room Humidity Test Data 34
9 Property 2 Living Room Light Test Data 34
10 Property 2 Kitchen Motion Test Data 35
11 The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property . . 36
12 The results of changing the epoch number of the model when used on the Temperature Bluetooth data from the first property 37
13 The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property . . 38
14 The results of changing the epoch number of the model when used on the Temperature Bluetooth data from the first property 38
15 Results for Property 1 Temperature 39
16 Results for Property 2 Temperature 39
17 Results for Property 1 Humidity 39
18 Results for Property 2 Humidity 40
19 Results for Property 1 Light 40
20 Results for Property 2 Light 40
21 Results for Property 1 Motion 40
22 Results for Property 2 Motion 40
23 System time taken for the model to process the Bluetooth temperature data for Property 1 41
24 Identified anomalies in Property 1 temperature data for different solid value thresholds 45
25 Identified anomalies in Property 1 humidity for different solid value thresholds 46
26 Table showing identified anomalies in Property 1 temperature for different percentage value thresholds 47
27 Table showing identified anomalies in Property 1 humidity for different percentage value thresholds 48

28 Results of utilising a combined approach of temperature with humidity . . . 56
29 All readings in property 1 temperature supplemented with humidity . . . 58
30 All readings in property 2 temperature supplemented with humidity . . . 58
31 Results of utilising a combined approach of temperature with light . . . 60
32 All readings in property 1 temperature supplemented with light 60
33 All readings in property 2 temperature supplemented with light 60
34 Results of utilising a combined approach of temperature with sound . . . 61
35 All readings in property 1 temperature supplemented with sound 61
36 All readings in property 2 temperature supplemented with sound 61
37 Results of utilising a combined approach of humidity with temperature . . . 62
38 All readings in property 1 humidity supplemented with temperature . . . 62
39 All readings in property 2 humidity supplemented with temperature . . . 63
40 Results of utilising a combined approach of humidity with light 64
41 All readings in property 1 humidity supplemented with light . 64
42 All readings in property 2 humidity supplemented with light . 64
43 Results of utilising a combined approach of humidity with sound 65
44 All readings in property 1 humidity supplemented with sound 65
45 All readings in property 2 humidity supplemented with sound 65
46 Table showing all readings in property 1 temperature 74
47 Table showing all readings in property 2 temperature 75
48 Table showing matching readings in property 1 temperature . 76
49 Table showing matching readings in property 2 temperature . 76
50 Table showing all readings in property 1 humidity 79
51 Table showing all readings in property 2 humidity 79
52 Table showing matching readings in property 1 humidity . . . 81
53 Table showing matching readings in property 2 humidity . . . 81
54 Table showing all readings in property 1 light 84
55 Table showing all readings in property 2 light 84
56 Table showing all readings in property 1 motion 87
57 Table showing all readings in property 2 motion 87

58 The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property with a 9 to 1 split 120
59 The results of changing the epochs of the model when used on the Temperature Bluetooth data from the first property with a 9 to 1 split 120

2 Acknowledgments

I would like to thank the following people for their assistance during this research project:

My supervisor Dr Alia Abdelmoty for providing guidance and assistance at every stage of the project.

KESS2 for providing funding that allowed me to undertake this project in the first place.

Everyone at the Safehouse team for providing data for us to use in this project as well as their regular feedback during the development of the system.

My family and partner for their patience and support, especially when I was working on my thesis from home.

3 Introduction

A growing problem across the United Kingdom is the emergence of an aging population. This problem is expected to grow, as the ONS (Science 2016) predicts that by 2050 over 25% of the UK's population will be 65 or over. (*Overview of the UK population - Office for National Statistics* n.d.)

As people get older they are more likely to require regular care. Currently, many families are taking to caring of their relatives in need. However, this can have negative effects on the carer as they may not be prepared to provide full-time care without negatively affecting their own quality of life due to the strain put on them, both monetary and mentally. (*Living longer - Office for National Statistics* n.d.)

Many elderly people are not able to be cared for by their relatives and instead need to receive professional care, either at home, or in a care home. While waiting for these care options to become available, many people are required to stay in hospital beds. Despite this, the number of hospital beds available for this purpose has decreased substantially over the last few years (Ewbank, Thompson, and McKenna 2017). This is a growing problem and without practical solutions to solve them will result in a decrease in the quality of life for many older people in the country.

In this study we will be introducing a system to detect anomalies in data from several different sensor sources, with the aim to answer the question of whether a machine learning system paired with environmental data can help to allow elderly and vulnerable people to live at home, without the need of constant care. We will be analysing and utilising sensor data from two properties provided by our partner company Safehouse.

Several different machine learning algorithms will be compared to identify which provides the most effective and reliable results for our stated goals. The algorithms tested will be identified through the use of relevant related works.

This study will be focusing on detecting anomalies. According to the Cambridge Dictionary an anomaly is defined as 'a person or thing that is different from what is usual, or not in agreement with something else and therefore not satisfactory'. For the sake of this study an anomaly is perceived as any value received from a sensor that is out of the ordinary range for that sensor in the specific time period. These points will then be evaluated by the experts at Safehouse to ensure that expected anomalies are being detected, and that potential false positives are ignored. With their help the system will be refined until a good degree of accuracy is achieved.

Machine learning was deemed necessary for this study due to the potential limitations of traditional data analysis and anomaly detection. Machine learning is good at working with large data sets when compared to an individual manually sorting through this data. (Zhou et al. 2017), (L. Wang, G. Wang, and Alexander 2015). In particular machine learning can be used in order to detect patterns and trends that would be missed by a non machine learning algorithm.

Prediction will be used in order to find what the expected value for the sensors will be at a specific time. This is important for obtaining a base value to compare the actual value of the sensors against when attempting to detect anomalies. The accuracy of the machine learning model used will dictate how accurate the expected value for these sensors will be at a given time, helping to ensure that the anomalies detected are genuine. While an average value for these sensors could be predicted without using machine learning, the use of machine learning should allow us to more accurately predict the future values for these sensors. This type of future prediction is regularly used in other markets, such as the stock market where it is used in order to decide when to buy and sell stocks. (Patel et al. 2015) (Shen, Jiang, and Zhang n.d.). This is achieved by monitoring the values of stocks over time and detecting any trends and inconsistencies. While this paper is focused on sensor data, rather than stock value, the concepts behind either methods are applicable in both situations, as we are aiming to detect differences in a single value over time.

One of the most important aspects of the design of the system is how well it is able to convey information to the user. To achieve this a bespoke user interface will be designed that allows for the user to choose and customise what information is presented, and how. The interface will be designed to allow the user to see at a glance which anomalies have been detected, as well as allowing them to gain more insight into why they occurred by displaying a timeline of events.

In this study we suggest a potential system that would help to allow elderly, and otherwise vulnerable, people to live alone. This will be through the implementation of an anomaly detection application that can act as an early warning system for carers. Allowing the vulnerable person to be left alone, unless the system detects there is a reason for the carer to intervene. To achieve this we aim to address several key research questions.

- Can environmental sensors be used in combination with automatic remote monitoring to allow vulnerable people to live alone?

- Can machine learning be used on environmental sensors in order to detect anomalies in human behaviour?
- Does a combination of different sensor types provide more accurate anomaly detection results than a single sensor solution?
- Can a system be designed to provide this information in an easy to read way?
- What is the most effective machine learning model for detecting anomalies in sensor data?

This thesis will be separated into sections in order to attempt to answer these questions.

INTRODUCTION: Introduction to the study, and the main motivations for doing this work.

METHODOLOGY: Details the concepts and techniques that were used in this study and provides background on some of the key terms used in this paper.

RELATED WORKS: Describes the solutions currently existing for similar problems, and what information is relevant and applicable to the research being done in this paper.

EXPERIMENT SETUP: Discussion of the problems that we aim to fix in detail. This section will also take into account the considerations that were made to ensure the study was fair and successful.

MODEL COMPARISON: Taking the models that were identified in the related works and testing them on the different properties and their sensors. Models will be compared to find out which performs the most effectively on our data. This will start with a comparison of the different configurations for each of the models. The ideal configurations of the models are then compared against each other on several different criteria in order to decide which model is the best under perfect circumstances.

ANOMALY DETECTION WITH SINGLE AND MULTIPLE SENSORS: Details what is considered an anomaly by the system. We will also discuss how the system is able to detects anomalies, as well as the results of using the system on the key sensors for both properties.

ANOMALY DETECTION WITH ADDITIONAL PARAMETERS: Implementation of the features suggested by our industry partners, and the testing results of combination sensors in the system. The results of each of these on the different sensors will be compared to determine in which cases these extra parameters are worthwhile.

USER INTERFACE: Covers the considerations taken to display data in a clear and concise way. We will also discuss how the feedback from our industry partner helped to shape the final design.

DISCUSSION: An evaluation and discussion of the work that has been done and the limitations that had an impact on the results of our study. Details of how the study could be improved given extra time and resources.

CONCLUSION: Final conclusions on whether the system was able to achieve the goals that were set out in this introduction.

4 Methodology

4.1 Background

This subsection will cover the knowledge needed in order to understand the rest of this paper. We assume that the reader has a basic understanding of programming and machine learning before beginning.

4.1.1 Machine Learning

This section will briefly explain some of the machine learning terms that are used in this study.

Time Series Prediction Time series prediction is the process of using historic data in order to predict future values. Time series prediction will be used to map the expected future values for the different sensors.

Univariate and Multivariate Univariate is a term that describes data that only has a single characteristic. An example of univariate data in this study is the readings that come from a single sensor, such as the temperature over time.

Multivariate differs from univariate as it data with multiple characteristics. We will be using this type of data when utilising combined data sources such as those coming from multiple sensor types.

ANN Artificial neural networks are designed to artificially simulate the neurons of a brain. They are made up of several nodes that are put together to match the structure of a brain. ANNs are made up of input and output layers, as well as several hidden layers that act to change the input into data that can be used by the output layer.

ANNs excel in situations where there is a large amount of data that needs to be monitored to find patterns. Through the use of ANNs complex patterns that would be impossible for single individuals to find through conventional means can be found.

There are several different types of ANNs, however for the sake of this study we will be looking at only one type of ANN known as a recurrent neural networks.

RNN Recurrent neural networks are a type of ANN with specific properties that make them particularly suited for learning and predicting on complex data.

Like other types of ANNs, RNNs take in training data to learn. This training data is a sample of a representative data set that is similar to the data the model will eventually be used on. In many cases this training data is made up of a portion of real world data.

Unlike other types of ANNs, RNNs have a temporal dimension, and as such take into account the order and sequence of data. This is important when trying to learn time series data, which is important for the purpose of this study.

4.2 Machine Learning Models

Several different machine learning methods will be tested in this paper. As will be discussed in the section 5.3 of this paper, these were the models that were successful in other papers with similar scenarios. The details on how these models work are provided below.

4.2.1 LSTM

LSTM, which stands for Long Short Term Memory, is a type of recurrent neural network that recognises patterns in data. Unlike other neural networks, LSTM takes into account the time and sequence of data, making it particularly suited for making forecasts on time series data.

This paper will primarily be looking at univariate LSTM models, however later tests will be performed using multiple sensors and will therefore need to use multivariate models. In a univariate model only a single variable is used as an input, here the variable would be the data from a single sensor. Multivariate models on the other hand takes multiple variables as an input.

The implementation in this study was created using the Keras library with a TensorFlow back end. This is the standard method for implementing machine learning models using the Python programming language. Both Keras and TensorFlow are the leading standard for implementing machine learning and are used by several industry leaders, such as Airbnb, Intel, Twitter and Google.

4.2.2 AutoEncoder

An AutoEncoder will be tested for its ability to make predictions and detect anomalies. This system works differently than the other time series machine learning methods as it's not attempting to predict the future, instead attempting to recreate the input provided based on past data. We will then be assessing how well the autoencoder's predictions matched up with the actual results coming from the real data from the sensors.

The autoencoder receives an input and attempts to learn and recreate that input as accurately as possible. To help with this process, LSTM layers are used by the model to capture the importance of the sequentiality of the data that is used as an input to the model. Using LSTM layers should dramatically increase the accuracy of the model in this instance.

4.2.3 ARIMA

Auto Regressive Integrated Moving Average, also known as ARIMA, is a generalisation of time series data that can be used to predict future values by using past data. It is generally accepted that any non random time series data can be modeled using ARIMA, with varying degrees of success depending on how much of a pattern is present within the data. While standard ARIMA is non seasonal, the results of the ARIMA model can be improved in data sets where there is some seasonality by utilising Seasonal Auto Regressive Integrated Moving Average (SARIMA).

ARIMA is made up of three terms, p, d and q. P is the lag order, d is the degree of differencing, and q is the order of moving average. Generally these three properties are adjusted in order to find the ideal performance for a given time series data set. However, we will be utilising a utility known as Auto ARIMA that automatically tests the different potential models to find the ideal fit for the data. For the sake of this study the Pyramid ARIMA library's implementation of Auto ARIMA will be used.

4.3 Measuring Model Accuracy

In order to ensure that the models being used are suitable for the different sensors they will first need to be evaluated. Several different methods have been used in order to assess the accuracy of these models in different situations. As we have access to experts in this field, we will also be using their advice and guidance to fine tune the models. The methods described here were chosen based on the papers found and discussed in the related works section 5.

4.3.1 MAE

Mean Absolute Error (MAE) is a measure of the error between the actual value and the value predicted by the model. It is the average of the absolute errors.

This method of measuring model accuracy is regularly used, as shown by our related works research in section 5.4, and can be applied on each of the prediction models. MAE is particularly well suited as a standard measurement method as it purely describes the average error. The issue with MAE is that it is using the absolute values which can be problematic in some cases, particularly when attempting to compare the results of multiple sensors.

4.3.2 RMSE

Root mean squared error (RMSE) is the average amount of error in the model's predictions in comparison to the actual values in the data set. In general, a lower RMSE value is considered better, as this indicates that the average amount of error in the model's predictions are low.

RMSE is dependent on the type of units being measured, therefore it can be difficult to compare the RMSE of models on different data sets. For example in temperature a few degrees change would be large, while in a sensor like light, a change of a few lux is not significant.

One method of evaluating the RMSE is to compare it with the standard deviation of the data. This will provide a baseline for how well the model is learning, in comparison to the average amount of difference between the different points in the data set. A model with a lower RMSE than the data sets standard deviation is learning from the data and performing positively, while a higher RMSE indicates a problem with the training model.

4.3.3 R2

R-squared, or R2, is a measurement that describes how much of the data variance can be explained by independent variables in the regression model. Unlike RMSE, R2 is not dependent on the type of unit being measured. For this reason it is possible to measure and compare multiple models against each other effectively by using the R2 score.

When measuring R2, a value closer to 1 is generally better. Due to the randomness of human behaviour, this may not always indicate that this is the best model however.

4.3.4 Expert Feedback

As a lot of the metrics used to measure model accuracy are subjective, many of the judgements will be made based on the expertise of our partner company, Safehouse. They have several years of experience in the field and have been able to advise us on how effective they believe the model has been at any given time.

Throughout the project weekly meetings with the Safehouse team have been held. During these meetings the progress that had been done on the model implementation and user interface was shown to the attending team members. Generally these team members would include the CEO of Safehouse, a product specialist and a software engineer. This range of experiences helped to ensure that the feedback provided was relevant to every section of the project. The product specialist was particularly knowledgeable about the sensors and their real world implementations. The feedback we received from the team was entirely qualitative. This is due to how subjective the measurement of how well the system performed was, requiring us to rely on the judgement of these experts.

Because we have been showing the user-facing side of the application, the team at Safehouse have had several opportunities to view the visualised data. Here they have been able to observe how the data has changed over time, and at which points exactly anomalies have been detected. This makes it easier for the team to give feedback on how the system has performed.

4.4 Model Comparisons

In the model comparison section we will be aiming to find out which of the identified models is best suited to the task of accurately detecting anomalies in the data set. To achieve this we will be measuring the accuracy of each of the models under different conditions and with different configurations. Once the ideal configurations have been found the 3 models will also be compared against each other under realistic conditions to decide on the best fit for the project.

Each of the models will be compared and their results analysed to discover their strengths and weaknesses. In order to ensure that the tests performed are fair to the different model architectures, preliminary steps are taken.

These measures will differ between the models as they all have different requirements for being effectively tested.

Before comparing models against each other the model architectures need to be tuned and optimised to provide the best possible results independently. To achieve this each of the models will go through several rounds of testing wherein their configurations will be changed and adjusted until the best model architecture is found for the test data set.

4.4.1 LSTM

Due to variance in results when training LSTM models, several repeats of the model training and testing are needed for this model to obtain an accurate idea of the results of using this model on the available data sets. In this study the combined average of 10 repeats will be used in order to assess this model. This will provide an average assessment for the error rates that the model provides when tested on the data set. A repeat in this instance is running the model on the same data with the same configurations multiple times.

Different configurations of this model will be tested in order to find the best suited for the data sets that we have. There are two key components of the LSTM configuration that can be modified to obtain better results, namely the batch size and number of epochs. The batch size is the number of training samples used in each pass through of the data during the training stage, while epochs are the number of times that the model passes over the entire training data set during training. Adjusting these parameters can have a large impact on the performance of the model. These configurations will then be measured using the methods outlined above, these are the RMSE and R2.

Once the model configuration has been decided upon the expert feedback will be taken into account to evaluate the effectiveness of the model. We choose not to involve them in the earliest stages as the different model configurations can be easily compared against each other with little worry that the results may not be representative of the models performance.

4.4.2 AutoEncoder

The AutoEncoder is the most simple to assess as no repeats are needed. As an LSTM layer is being implemented, different configurations of this setting will be tested to find the best AutoEncoder model for the data sets. These configurations will be assessed in the same way as the standard LSTM model, with feedback from our experts once the final configuration has been selected and testing done.

4.4.3 ARIMA

Auto ARIMA will automatically select the best model based on the AIC and BIC scores of each possible model configuration. For this reason experimentation with different ARIMA configurations is not necessary. Preliminary testing showed that seasonal ARIMA performed better than non seasonal ARIMA by a significant degree. Therefore all of the tests performed in this study will be using a seasonal auto ARIMA model.

4.5 Anomaly Detection

Once the models have been sufficiently trained we will then use a threshold system in order to detect anomalies in the data of each of the sensors. This will be done by first comparing the expected values provided by our predictions and comparing them against the actual values from the sensors. These difference between these values will then be compared with the threshold system to decide whether they should be detected as an anomaly, and how severe they are as an anomaly.

We will primarily be relying on the expert feedback from our partners in the Safehouse team in order to judge how successful our anomaly detection system is. This consists of regular meetings where the team can view the system and any anomalies that have been categorised. Using their expert knowledge they can then judge whether all of the expected anomalies have been detected, and how to tune the model further in order to get the desired results for later meetings.

4.6 User Interface

The interface of our system is also an important factor to consider. As with the anomaly detection we will primarily rely upon the feedback provided by the team at Safehouse. During the regularly scheduled meetings the user interface, including prototypes, will be shown to the team to gather feedback. The interface will then be molded around the feedback that they provide in each session in order to ensure that the system conveys information in an effective and fluent manner.

5 Related Works

We will discuss the relevant works that have been conducted on similar questions as our own. These will be split into sections depending on what material from these papers are relevant to our study.

5.1 Similar Papers

While there exist several other similar papers to this study, there are several key differences that differentiate this work from others. One of the most unique factors of this study is the type of data that is being collected, and the hands off approach that was taken when gathering this data. A unique range of sensors was installed within the properties, providing us with a variety of different environmental data. This differs significantly from other papers that have primarily focused on data received directly from the participant either through wearables (Zhu, Sheng, and Liu 2015) or through monitoring their actions directly using cameras (Deep et al. 2020) or RFID (Hsu and Chen 2010). This has a significant limitation in that the person is always aware that they are being monitored and may need to make alterations to their lifestyle to accommodate these extra sensors. Our approach to data gathering aims to be unobtrusive and considerate for the people living within the properties. This has already proved to be an important factor for technology acceptance by older people (Novák, Biňas, and Jakab 2012).

Unlike other papers, such as (Aran et al. 2016), we are not focusing on anomaly classification. This was not possible due to the unobtrusive nature of this study as in order to classify sensor readings we would need to keep a detailed diary of the properties being monitored. Instead we are choosing to simply focus on identifying potential anomalies in the data, leaving the classification to the experts later on.

Other studies in this field have chosen to hone in on participants with specific illnesses, such as (Arifoglu and Bouchachia 2017) and (Lotfi et al. 2012) that chose to focus on participants suffering with dementia, allowing them to tailor their study to focus on detecting known behaviours. One such behaviour (Arifoglu and Bouchachia 2017) chose to look for was the participant "forgetting and repeating activities". This allowed for those studies to be tailored to find a specific problem. This is not the case in this study as we are not focusing on a specific debilitating condition, instead we aim for this system to be applicable to all elderly or vulnerable persons.

One of the things that wasn't present in any of the studies currently available is how information is presented to a caregiver once an anomaly is detected. We believe that this is an important step of the process as information needs to be clearly conveyed to a carer for them to react. For this reason we will be discussing this in section 10 of this paper.

5.2 Living Alone

There are many challenges that the elderly and otherwise vulnerable people face due to living alone. While not all of these people require a constant carer, they may still feel the impact of being alone for long periods. These can include increased feelings of isolation, depression and suicidal thoughts as shown by (Fukunaga et al. 2012), (Dean et al. 2016) and (Yeh and Lo 2004). In the event that they are able to obtain a carer this can regularly be in the form of a family member who may not always be equipped to deal with caring for a vulnerable person. This can have a negative impact on both their, and the person receiving care's quality of life. (*Living longer - Office for National Statistics* n.d.)

5.3 Time Series Forecasting

In past studies different algorithms have been used to create future forecasts of time series data. Of those algorithms machine learning models performed particularly well for this task, with various Long Short Term Memory(LSTM) and Auto Regressive Integrated Moving Average(ARIMA) variations proving particularly popular and effective. Comparisons of these algorithms for time series prediction have had mixed results, with algorithms performing well in specific circumstances, but failing in others. This is generally due to several key factors, such as the type of data used in the study, or the amount of training data available.

A comparison of ARIMA and LSTM by (Siami-Namini, Tavakoli, and Siami Namin 2019) found that "deep learning-based algorithms such as LSTM outperform traditional-based algorithms such as ARIMA model." With an average error rate reduction of between 84% and 87% when compared to a standard ARIMA algorithm. Indicating that the LSTM algorithm was vastly superior to ARIMA when forecasting future data. This was contradicted however by a study attempting to predict future Bitcoin prices (Yamak, Yu-jian, and Gadosey 2019), as they found that ARIMA gave better results than the two deep learning regression models that they tested. They also found that another model, GRU, outperformed LSTM in their tests. While not directly comparable to our study, predicting the future value of bitcoins are similar to sensor readings as they are both individual numerical values that change over time.

In a study by (Siami-Namini, Tavakoli, and Siami Namin 2019) they found that changing the number of epochs in the LSTM model did not affect the model results in a significant way, instead it was found that the model forecasts exhibited a “truly random behaviour”. This indicates that attempting to tune the model for more accurate forecasts has no bearing on the actual effectiveness of the system. Instead more importance should be placed in training the model on a representative and sufficiently large subset of the complete data set.

Other comparisons with ARIMA have also been made. (Kohzadi et al. 1996) compared ARIMA with a feed forward neural network for forecasting the commodity prices of cattle and wheat between 1950 and 1990. They found that the neural network was able to perform significantly better than the ARIMA equivalent in all measurement categories used, namely MSE, MAE and MAPE. It was also found that while ARIMA was only able to detect sharp changes in the wheat data, the neural network was able to detect changes in both wheat and cattle, suggesting that the model is ”not problem specific” and would also be effective for predicting other data types, such as stock prices.

The study by (Khashei and Bijari 2011) attempted to use a hybrid approach of combining ARIMA with artificial neural networks in order to overcome the usual limitations of ANNs and to increase the accuracy of their own predictions. On the three readily available data sets that they tested on, it was found that the proposed hybrid model worked effectively and was able to enhance the results found in comparison to using components of the model separately. While (Hosseini and Sarrafzadeh 2019) used clustering, along side an LSTM autoecoder, to predict negative health events before they happened. They were able to predict anomalies in the health of infants with relatively high success, and suggested that their approach could be viable for identifying health events in other instances.

While the majority of the time series studies here focused on univariate data, some, such as (Sagheer and Kotb 2019) opted to use multivariate data for their predictions. This suggested method proved to perform better than alternative univariatie models tested in other papers.

From the research done here we can see that no single model solution performed better than the others in every instance. Of the papers tested both LSTM and ARIMA appeared to be particularly effective for time series prediction, for this reason these will be the models we look at going forward. While the majority of work done was using univariate data, the paper by (Sagheer and Kotb 2019) indicated that multivariate solutions are a worthwhile method to be explored.

5.4 Model Accuracy

Several different methods have proven effective for measuring the accuracy of the models once they have been implemented. Popular methods include Root Mean Square Error(RMSE), Mean Absolute Error(MAE), Mean Absolute Percentage Error(MAPE), Mean Absolute Squared Error(MASE), Root Mean Absolute Error(RMAE) and Coefficient of Determination(R2). For the purpose of this study we will be using the methods that prove to perform consistently in similar use cases as our own.

MAE proved to be one of the most popular methods of measuring model accuracy. However (Willmott and Matsuura 2005) found that MAE was an unambiguous and natural measure of average error. In comparison they found that RMSE was highly ineffective as a measurement metric because they found "no consistent functional relationship between RMSE and average error" and that there is no clear interpretation of it as a measurement. On the other hand (Chai and Draxler 2014) suggested that the RMSE is not ambiguous in meaning, contradicting several other papers that came before it. They found that when the error distribution for a model is expected to be Gaussian that RMSE was a more appropriate method of evaluating performance that MAE, despite the fact that many other researchers appear to favour MAE over RMSE. Overall they found that while RMSE is a more accurate measure of performance under certain circumstances, that a combined approach of multiple different measurement metrics are needed to get an accurate estimate of a model's performance.

Alternatives to using the standard MAE were also suggested. One such metric used in several papers was the MAPE, or Mean Absolute Percentage Error. Unlike MAE this provides a relative estimate of the amount of error between the actual value and the prediction, (Myttenaere et al. 2016) found that this was a feasible measurement method.

Another suggestions was the use of a dynamic mean absolute error, DMAE, as a measurement system. As suggested by (Frías-Paredes et al. 2018) this dynamic solution would attempt to account for both temporal and absolute error components. They found that their dynamic solution, paired with the standard MAE, provided a deeper and more accurate estimation of the performance of a models ability to estimate future time series data.

The paper by (Hyndman and Koehler 2006) compared several different forecast measurement methods. Their study suggested that scaled errors should become the standard measure for forecast accuracy. Particularly in cases where the data is of a very different scale they recommended the use of MASE, mean absolute squared error, to measure the accuracy of the forecast. They did however recognize that both MAE and MAPE can be useful for data on a smaller scale as they are significantly easier to explain.

As explored in these research papers we can see that each of these accuracy measurement methods have their own advantages, and disadvantages. As no clear consensus was found on the best measurement system we will focusing on a combined approach using RMSE, MAE and R2.

5.5 Anomaly Detection

Anomaly detection can be implemented in several ways. The method used is dictated by what is classified as an anomaly, as well as the type of data that is being predicted.

One of the main methods of anomaly detection is to look for outliers in the data, that is points that don't match the general pattern of the rest of the data set. The most obvious outliers are usually the highest and lowest points in the dataset, however through the use of machine learning, among other techniques, outliers that do not appear to deviate as much can be detected. In time series data these may be points that are vastly different than usual for a specific time of day.

The paper by (Kieu et al. 2019) recommended an ensemble auto encoder solution in order to reduce the problems associated with using a single auto encoder, primarily the issue of the autoencoders being overfitted to outliers.

5.6 Sensors

In this study a range of different sensors will be used. A suite of custom environmental sensors provided by our partner company Safehouse will be the main focus of the study. While this exact configuration of sensors has not been seen in other studies, there are multiple cases of similar set ups being used and analysed in different scenarios, with some studies focusing on similar problems requiring time series prediction, as well as anomaly detection.

Our study required us to use properties with custom sensors installed, however other studies have used existing infrastructure. (Yassine, Singh, and Alamri 2017) chose to utilise the readings of already present smart meters to detect activity, while (Tsukiyama 2015) focused on whether the participant was using a normal amount of water. While other papers have chosen to install their own unique sensors, such as (Nishida et al. 2016) which attempted to detect how well elderly people are able to get around their house through the use of a handrail grip sensor.

In the study by (Van Kasteren, Englebienne, and Kröse 2010), they used a comprehensive suite of environmental sensors connected to a single network node. Temperature sensors, reed switches, pressure pads and infrared were used in order to monitor the participants activity in the home. The use of wireless, battery powered, sensors allowed them to conveniently install the sensors around the property in unobtrusive locations.

While we are not intending to monitor the participants location, several studies including (Ahmad and Mohan 2009), (Hsu and Chen 2010), (Wu 2012) opted to use RFID tags in order to find out where the participant was moving around the house. The majority of these studies focused on the user carrying around an RFID tag, or sensor, and measuring which other tags they came into contact with. They found that the RFID tags were small enough to be unobtrusive to the participants, resulting in no resistance from the participant.

6 Experiment Setup

6.1 Description

The primary purpose of this study is to design and implement a system that utilises machine learning in order to detect anomalies in the behaviour of elderly or otherwise residents. To create the most effective system three different models will be tested. LSTM and ARIMA models will be used for time series prediction, while an LSTM auto encoder will be used in order to recreate the time series data that is provided to it. The advantages and disadvantages of each of these three will be compared against each other in order to decide which is the most suitable for the task at hand. Refinement of these results will be achieved by implementing extra parameters. This will cut down on the number of potential false positives and give further insight into why they occurred.

6.2 Partner Company

Safehouse are a company that specialise in providing sensors for the purpose of monitoring properties, particularly those occupied by vulnerable people. They already have several properties setup with various different sensor types, and they will be providing the data used in this study. Several members of the team assisted us with this project. This included the CEO and other members of the team, such as the project specialist and software engineer, who have extensive technical knowledge, and experience with using the sensors that have been provided.

6.3 Setup

Two properties are being monitored for the purpose of this research. These properties have several different sensors installed, with a wide range of environmental monitoring abilities. Three different types of environmental sensors have been installed in the two properties. Each of these sensors have different purposes, and will provide relevant information for the rooms that they are based within.

This study will be using a mixture of main and battery powered wireless sensors. The main benefit of the battery powered sensors over mains powered is that they can be installed in unobtrusive locations, such as hidden on walls.

These low powered devices will also be able to last the duration of the study without requiring a battery change.

6.4 Environmental Sensors

6.4.1 Safehouse Hub

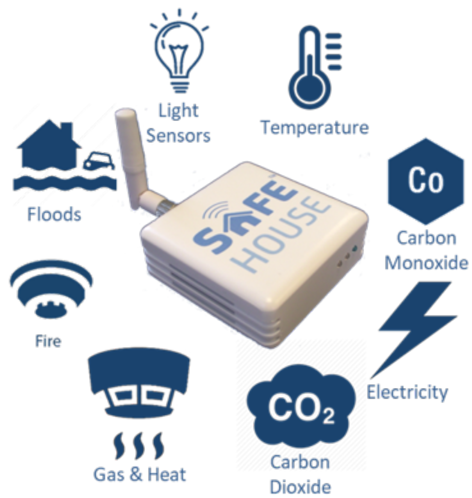


Figure 1: Safehouse hub

The Safehouse hub sensor acts as the main access point to the network. This sensor's primary purpose is to report the results of the other sensors back to the Safehouse database to be collected for this study. This is a mains powered sensor that will be located in a central location in the properties.

While this sensor is able to monitor some basic readings, because of the improved accuracy of the USB and motion sensors, we will be using those sensors where possible, with the hub acting as a backup in the event that one of those other sensors fails or runs out of battery.

6.4.2 Safehouse USB

Figure 2: Safehouse USB

The primary sensor being used in the two properties living rooms is the Safehouse USB sensor. This is the smallest of the sensors that will be used, but is also the most accurate in the data that it gathers. Like the hub, this sensor needs to be plugged directly into the properties mains, and will need to be placed in a convenient location for the user.

One benefit of this sensor is its storage capacity. The USB sensor is able to store several months worth of data internally, reducing the risk of losing data due to a power outage, or in the event that the main hub is not able to deliver data back to the main hub for any reason. Each of the sensors associated with this device transmit their readings every 30 minutes.

Sensor	Measurement	Purpose
Temperature	Degrees	Temperature comfort of the room
Humidity	Percentage	Humidity comfort of the room
Light	Lux	How bright the room is
Air-Pressure	Pascals	Pressure in the room
Noise	Decibels	How much noise has been recorded
VOC	Parts Per Million (ppm)	VOC particles in the air
ECO2	ppm	CO2 levels in the room.

Table 1: Table of the data tracked by the Safehouse USB Sensor

6.4.3 PIR Sensor

Figure 3: Safehouse PIR

We will also be using a more specialised sensor for our study, the passive infrared (PIR) sensor. This sensor works in the same way to the basic hub sensor, with several extra features that are available due to the PIR camera that has been built into this device. The unique benefit of the PIR is the ability to track motion within its range of sight.

The PIR sensor works by measuring infrared light emitting from objects in its proximity. PIR sensors give a general idea of how many times an object has passed by it, but is unable to detect what that image may be. For that reason this sensor can only be used to track the number of times that an occupant has moved past the sensor. The device will record the number of times that the PIR sensor detected movement within the space of an hour and report that back with the rest of the results.

One of the limitations of the PIR sensor is that it is unable to differentiate between the different things that may activate it. In houses with multiple occupants, or pets, this may cause issues as each of them would trigger the sensor. While we cannot control whether guests trigger the sensor, to reduce the chance of pets activating the sensor they will be placed at the occupants chest height, allowing animals to walk below the beam without setting it off. For the sake of this study we will be assuming that only the occupant of each property is triggering the PIR sensor.

This device is battery powered and unlike the other sensors used is not directly linked to the mains supply. However because the battery is tested to last several years on low power mode it will not run out during the duration of this study. The main advantage of this is that the sensor is not limited in install location by where it can be plugged in, particularly important as these sensors need to be located where they're most likely to detect only the occupants movement.

The Safehouse team opted to install these sensors at chest level and near the entrance of the rooms. This would ensure that the sensors detect the occupant as they move into and leave a room, while remaining out of sight.

Time Frame	Hourly
Temperature	Degrees
Humidity	Percentage
Light	Lux
Air-Pressure	Pascals

Table 2: Table of the data tracked by the Safehouse PIR Sensor

6.5 Data Reporting

Data is gathered from the environmental sensors at regular intervals where they report all data back to the Safehouse central database. This reporting is kept consistent as all of these sensors are kept in a static location and are connected to a local network through a central hub sensor. In the event that the hub sensor is unable to transmit for any reason, a notification is provided to the Safehouse team who can then work to resolve the issue.

The environmental hub sensors contained within the properties report back to the main database once every hour. However the separate environmental sensors themselves report at different intervals, ranging from every 20 minutes to hourly.

A method of preventing data loss is also present, as many of the sensors are outfitted with local storage systems that can be used to recover data that would normally be lost in the event of them failing to transmit.

Data is pulled from the central Safehouse server every hour through the use of an API. Once processed, this data is then used to update a Mongo database stored on Mongo Atlas. Mongodb was chosen as the database storage method as it is highly scalable and can handle the ever increasing amount of data that is passed to it. The database is hosted on the cloud through Atlas with backups hosted on multiple server providers, ensuring that the database has a high rate of availability.

NoSQL was chosen as the database as it allows for the type of data being stored to evolve and change over time. Flexible model updates are one of the key benefits of NoSQL databases. Allowing for different fields to be added, or removed, from within different documents in the same collection. This was particularly important in this instance as different rooms had different sensor types.

There were also instances early on in the project where the Safehouse team would install new sensors into the properties, which would require whole model changes in a regular SQL relational database, but was not necessary here.

To facilitate the inclusion of sunrise and sunset data into the system the 'sunrise-sunset.org' API is being used. This provides the sunrise and sunset data for each day in the participants location. This level of accuracy helps to determine which anomalies detected were in the day and which were in the night, which can be important for assessing the severity of certain anomalies.

6.6 Participants

This study is primarily focusing on users of a similar age range who live alone, but no specific health conditions are being looked for. For this reason the specific conditions seen elsewhere such as (Arifoglu and Bouchachia 2017),(Lotfi et al. 2012),(Arifoglu and Bouchachia 2017) were not necessary for our study. Instead we will be monitoring the properties of two elderly patients with no significant health conditions.

6.7 Property Map

The layout of the property being monitored, as well as the sensor placements appeared to have a drastic effect on the quality of data. (Aran et al. 2016) found that they had much more accurate results by making sure that their sensors were working correctly, and that they were placed in appropriate locations in the property. (Rabiner and Juang 1986) corroborated this as they found that the layout of houses had a large impact on the system's results, and that using raw sensor data could result in inconsistent results

An aim of the sensors is to be as unobtrusive to the occupant as possible. To achieve this the sensors have been carefully placed in locations where they won't be in clear sight, while ensuring that they're still in the optimal location to accurately detecting their environment.

The exact placements of the sensors was decided by the team at Safehouse. Their expertise allowing them to place the sensors in the locations that would provide the most accurate data, while remaining out of the occupants sight.

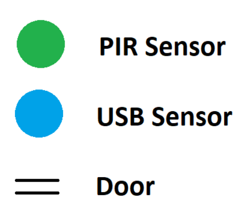


Figure 4: Floor Plan Key

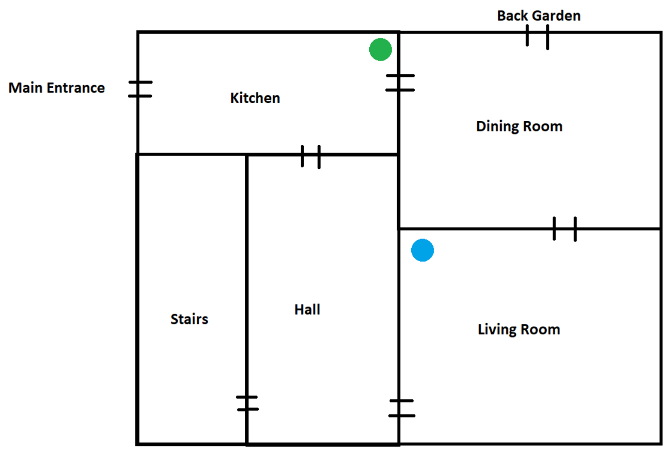


Figure 5: Property 1 Ground Floor plan

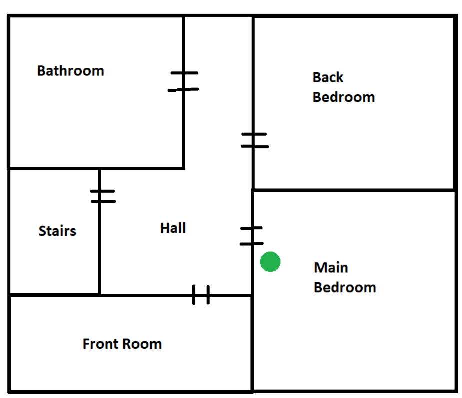


Figure 6: Property 1 First Floor plan

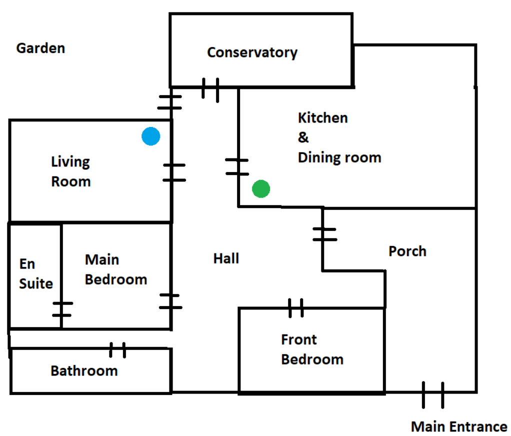


Figure 7: Property 2 Ground Floor plan

7 Model Comparison

In this section we will be displaying the results of utilising the various machine learning models on the data set that is available. We will first test the differences that different configurations have on the models, and then we will use these tuned models on the individual sensors.

7.1 Model Consideration

Several methods for detecting anomalies in the data sets have been considered. As no clear consensus was achieved as to the ideal model in the related works section 5.3, a mixture of model types will be used to identify the best fit for our situation. The three most prominent methods identified for time series prediction were LSTM, ARIMA and AutoEncoder. Each of these models have their own distinct advantages and disadvantages that will be explored and assessed.

These models were selected due to their prevalence in similar works, as discussed in section 5.1 and 5.3. Past studies (Siami-Namini, Tavakoli, and Siami Namin 2019), parenciteYamak2019AForecasting, (Kohzadi et al. 1996) have shown the effectiveness of LSTM and ARIMA for predicting future time series data. ARIMA also provides a good base line performance that the models will need to be able to exceed to demonstrate their effectiveness.

As this study will be attempting to detect anomalies in the data an AutoEncoder solution is also being tested. This model will include an LSTM layer, and as such should receive some of the advantages of using LSTM to forecast future data.

7.2 Data

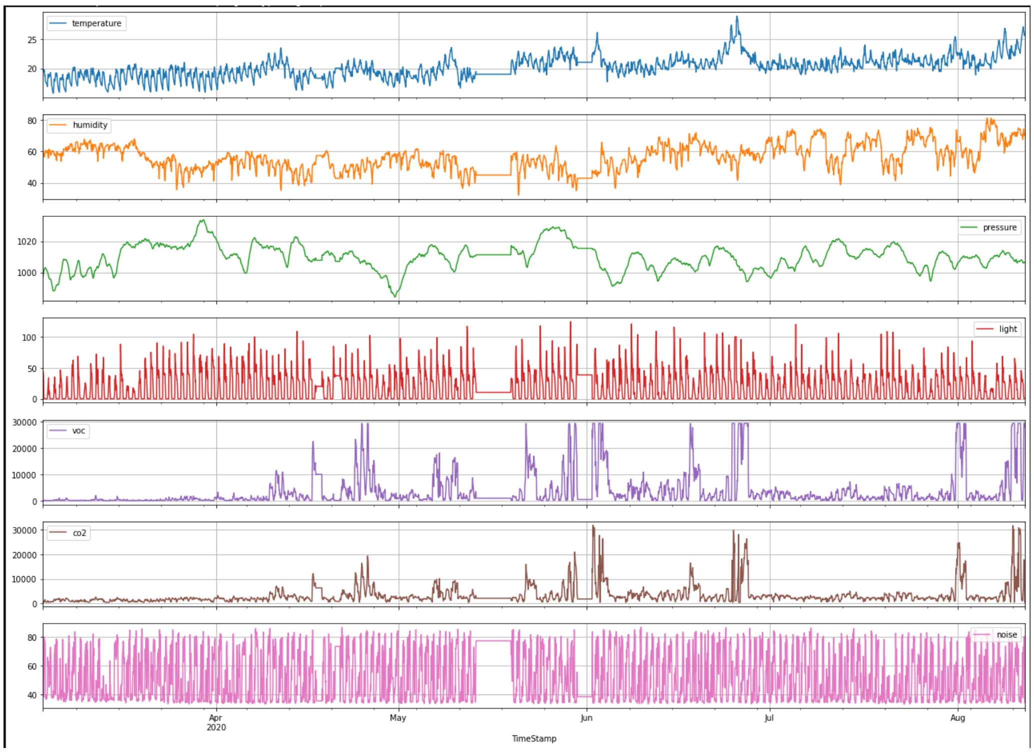


Figure 8: Property 1 Bluetooth data

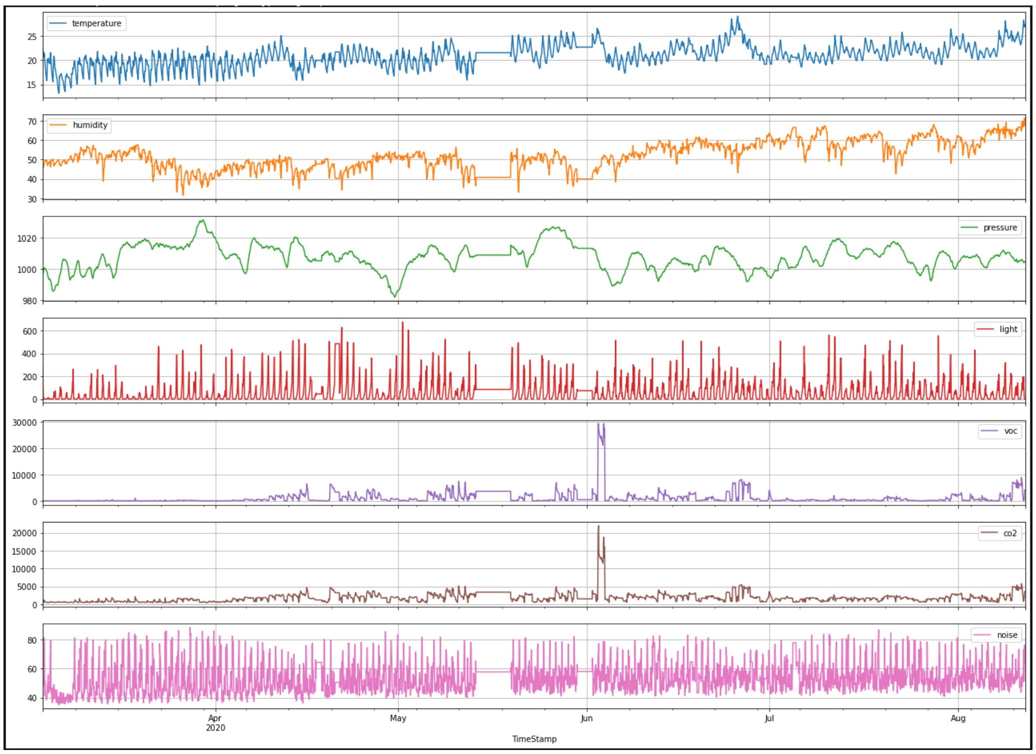


Figure 9: Property 2 Bluetooth data

The data consists of several anonymous data sets from two different properties. These data sets have been provided by the team at Safehouse and are stored in a secure MongoDB database.

While the data sets for the two properties were similar, they were not identical due to limitations related to data gathering early in the project, particularly due to the difficulty in procuring the required sensors at the beginning of the project. This led to a delay in sensor installation and the eventual decision to install one less PIR sensor in property 2.

Where possible the Bluetooth sensors in the living rooms were used for testing the system, as these had the highest accuracy for the data that we were recording. Before processing both property 1 and 2 had Bluetooth living room data records between 03/03/2020 and 11/08/2020. Property 1 consisted of 27125 records, while property 2 consisted of 27111. As we wanted to test the performance of the system on motion data we needed to utilise a sensor with motion capturing capabilities. To achieve this we used the PIR sensors located in the kitchens of the two properties.

To keep the data consistent, prior to use in the system the data is first re-sampled based on computing the mean for a given time frame. While this is customisable in the interface, the default is to re-sample to the hour. An hour was chosen as the default value as this would give a good balance between the number of data points available while remaining a short enough time frame to remain actionable by a carer.

In the event of gaps in the data set these are filled in with the nearest following value. This is ideal for sensors such as the temperature and humidity as the variation between two hours should not be large. As can be seen in the figures 8 and 9 there was one notable gap in the data where this was needed. This was between the months of May and June where the sensors were not able to transmit data.

Before being used by the model, each of the data sets will be scaled using the sklearn StandardScaler. This standardised features by removing the mean and scaling to unit variance. Scaling helps to prevent larger values in the data set from having too big of an impact on the training of the model. Scaling can also be used to allow different features to be compared to each other, which is useful in multivariate prediction.

Once normalised the data is then split into training and test data. Due to the limited time of the data capturing period there was only a few months worth of data to train on. Preliminary testing using a ratio of 9:1 resulted in there not being a lot of test data to evaluate the models sufficiently. Particularly as this meant that there was less than a month worth of data available for testing the model. The results of these preliminary tests can be seen in 58 and 59. For that reason in our primary tests we use 70% of the data set for training, while the remaining 30% is used for testing, ensuring that there is enough data to use several months worth for training and testing the model before it is and applied on the remaining data.

The split on the data was made based on time without random resampling. This was done due to the importance of the data remaining sequential. The first 70% of the dataset was used for training, while the remaining 30% was used for testing.

As there are two distinct properties, different models will be trained on the different data sets available for each of them. This will occur at runtime of the application to ensure that the results are relevant and kept up to date. The configuration for these models will be kept the same between properties.

Validation was considered in order to evaluate the performance of the model, however due to the lack of access to participants it was not possible to record data for this purpose. This meant that there was no data to compare the results against, meaning we needed to rely on the measures detailed above in order to evaluate the system.

Statistic	Value
Count	3875
Mean	20.195
Standard Deviation	1.804
Min	15.753
25%	18.993
50%	20.135
75%	21.200
Max	28.897

Table 3: Property 1 Living Room Temperature Test Data

Statistic	Value
Count	3875
Mean	56.240
Standard Deviation	8.436
Min	32.120
25%	50.203
50%	56.133
75%	61.900
Max	81.383

Table 4: Property 1 Living Room Humidity Test Data

Statistic	Value
Count	3875
Mean	19.148
Standard Deviation	23.216
Min	0
25%	0
50%	10.333
75%	34.500
Max	124.333

Table 5: Property 1 Living Room Light Test Data

Statistic	Value
Count	4548
Mean	11.924
Standard Deviation	13.920
Min	0
25%	0
50%	8
75%	18
Max	113

Table 6: Property 1 Kitchen Motion Test Data

Statistic	Value
Count	3873
Mean	20.942
Standard Deviation	2.230
Min	13.195
25%	19.665
50%	20.970
75%	22.260
Max	29.050

Table 7: Property 2 Living Room Temperature Test Data

Statistic	Value
Count	3873
Mean	52
Standard Deviation	7.258
Min	31.65
25%	46.933
50%	51.68
75%	57.41
Max	71.43

Table 8: Property 2 Living Room Humidity Test Data

Statistic	Value
Count	3873
Mean	60.658
Standard Deviation	97.005
Min	0
25%	0
50%	17.667
75%	84
Max	678.333

Table 9: Property 2 Living Room Light Test Data

Statistic	Value
Count	3854
Mean	5.632
Standard Deviation	8.694
Min	0
25%	0
50%	1
75%	8
Max	74

Table 10: Property 2 Kitchen Motion Test Data

7.3 Configurations

The first step for deciding on the ideal model for the data set is to compare different configurations of the model types in order to find the architectures that provide the most accurate results. Different models will need to be modified in several ways to obtain more accurate results. This report will first be looking at optimising the models used before comparing them against each other.

During the testing process of the models the same data set will be used. This is to ensure that the results are kept consistent, and that differences in the data sets don't have a positive, or negative, effect on the model's accuracy. For example a dataset without much variation over time is more likely to return more accurate predictions when used by the model, whereas a dataset with a lot of variation would have the opposite effect.

For this reason the tests were performed on the temperature data from the living room Bluetooth sensor in the first property. This data was chosen because it was the largest chunk of data that was available early in the process that had no noticeable gaps. Temperature was chosen as the test sensor because it was clear that patterns could be identified in this type of data, due to the lack of variance in the data, and the peaks and lows remaining in similar locations each day.

7.3.1 LSTM

Batch Size	RMSE	R2
64	0.726	0.677
128	0.670	0.722
256	0.608	0.774
500	1.422	-0.264

Table 11: The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property

We can use the results of these experiments to decide which of the model configurations is the best fit for the data set provided. In this instance we can see that the batch size of 256 provided the lowest RMSE, and R2 closest to 1, for these reasons it is indisputably the best batch size of those tested.

Epoch Number	RMSE	R2
10	0.606	0.774
50	0.523	0.834
75	0.923	0.473
100	0.679	0.708

Table 12: The results of changing the epoch number of the model when used on the Temperature Bluetooth data from the first property

Similar results were found in the epoch number, where 50 provided the best results in both categories. In general as the RMSE value became higher the R2 score for the results were further away from 1. This is because the results are highly dependent on each other and a higher RMSE would indicate that the R2 would be worse when repeatably testing the same dataset.

From these experiments we can see that the ideal batch size is 256 and number of epochs is 50 for the data set that this model was tested on. Further tuning at a granular level may provide better results on specific data sets, but for the sake of future tests we will be going with the configuration found here.

7.3.2 LSTM AutoEncoder

Batch Size	RMSE	R2
64	0.123	0.083
128	0.122	0.106
256	0.140	-0.193
500	1.142	-0.215

Table 13: The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property

Epoch Number	RMSE	R2
10	0.134	-0.087
50	0.122	0.106
75	0.122	0.103
100	0.131	-0.031

Table 14: The results of changing the epoch number of the model when used on the Temperature Bluetooth data from the first property

From these tests we can see that the ideal configuration for the Autoencoder is a batch size of 128 and 50 epochs. As these are the two configuration options with the best relative RMSE and R2 values.

7.4 Comparison of models

In this section the models will be compared. Their performance will be measured using their MAE, RMSE and R2 as mentioned above. The amount of time that it takes for the models to fully compute will also be taken into account for these tests, as it's important that the system can be run in a timely manner. Timings will be measured using Python's built-in process_time method. This will measure the amount of time that the processor spends on each of the tasks, while ignoring time spent unrelated to the process. This should provide the most accurate timings for running the system.

7.4.1 Individual Sensor Results

Model Type	MAE	RMSE	R2
AutoEncoder	0.544	0.722	0.060
LSTM	0.176	0.242	0.915
ARIMA	0.610	0.881	-0.208

Table 15: Results for Property 1 Temperature

Model Type	AME	RMSE	R2
AutoEncoder	0.540	0.692	0.056
LSTM	0.083	0.126	0.973
ARIMA	0.566	0.748	-0.009

Table 16: Results for Property 2 Temperature

Model Type	MAE	RMSE	R2
AutoEncoder	0.636	0.858	0.420
LSTM	0.173	0.260	0.947
ARIMA	1.520	1.796	-1.509

Table 17: Results for Property 1 Humidity

Model Type	MAE	RMSE	R2
AutoEncoder	0.362	0.512	0.463
LSTM	0.106	0.176	0.946
ARIMA	0.567	0.737	-0.061

Table 18: Results for Property 2 Humidity

Model Type	MAE	RMSE	R2
AutoEncoder	0.856	1.212	-0.800
LSTM	0.203	0.388	0.815
ARIMA	0.753	0.894	0.004

Table 19: Results for Property 1 Light

Model Type	MAE	RMSE	R2
AutoEncoder	0.819	1.136	-0.664
LSTM	0.257	0.553	0.622
ARIMA	4.219	4.336	-19.455

Table 20: Results for Property 2 Light

Model Type	MAE	RMSE	R2
AutoEncoder	0.597	0.820	-0.518
LSTM	0.706	0.843	-1.374
ARIMA	0.661	0.756	-0.299

Table 21: Results for Property 1 Motion

Model Type	MAE	RMSE	R2
AutoEncoder	0.548	0.902	-0.115
LSTM	0.321	0.583	0.577
ARIMA	0.716	0.895	-0.071

Table 22: Results for Property 2 Motion

7.4.2 Time Taken

Model Type	Seconds
AutoEncoder	145.953
LSTM	12095.469
ARIMA	5136.531

Table 23: System time taken for the model to process the Bluetooth temperature data for Property 1

The one area that the models differed the most was in the amount of time it took for the system to process the data. Monitoring only the time that the process was being run on the CPU showed that the auto encoder was able to process significantly faster than the other two models.

ARIMA took almost thirty five times longer to run than the Autoencoder did. While LSTM took even longer, taking more than double the amount of time that it took for the ARIMA system to finish. This is likely because the LSTM model needed to be retrained on the new data to remain relevant each time.

These tests were performed on a top of the line consumer grade CPU, an AMD Ryzen 9 5950x. In a realistic scenario this system would need to be run continuously on a less powerful server, which would only serve to increase the length of the already long run times for the system.

Making the user wait for long periods of time does not provide a good user experience, and may also result in the user being unable to react quickly to detected anomalies in cases where the system took to long to finish processing.

7.5 Conclusion

From the tests done we can see that the standard LSTM time series prediction model demonstrated the most accurate results when it came to making predictions using the dataset, followed by the Autoencoder and with Arima proving the least effective. While the Autoencoder model was significantly quicker than either of the alternatives in terms of processing time.

The length of time that it took for the LSTM system to finish processing the data made it an unviable solution for use in a real world scenario, due to the expectation that this system would be running on low cost servers with less powerful hardware than the computer that is being used for testing. This issue could potentially be circumvented by the use of a powerful cloud server provided as part of AWS or Microsoft Azure, however this would likely incur an increased cost which may not be palatable for the potential users of this system. For these reasons the Autoencoder system is used going forward.

8 Anomaly Detection With Single and Multiple Sensors

In this section we will be detailing how the system was designed to detect anomalies within the data sets, as well as the results of running the system on each of the different data types.

8.1 Detecting Anomalies

Prior to being used on a sensor's data set the model is first trained on that individual sensor's training data. Once trained the model can then be applied on the sensors remaining data. This will provide a list of data predictions based on the data that was fed in.

The output is evaluated by comparing the models predicted values, against the actual values of the sensor. The error of this comparison is then compared with thresholds in order to decide whether a point is classified as an anomaly or not. Any values that exceed these thresholds will be classed as anomalies.

8.2 Thresholds

To decide what the ordinary range should be for a given sensor a threshold system will be implemented. This threshold will define how different the actual value is able to be from the predicted value before it is detected as an anomaly. Two different thresholds are used to decide on the level of severity for the anomalies that are detected.

If the error rate exceeds the low threshold, but not the high threshold, the point is considered a low anomaly, while a point that exceeds the high threshold is considered a high anomaly. A point that does not exceed either of the thresholds is not considered an anomaly.

Tests were performed with different threshold levels in order to evaluate which would provide the most effective results. These were performed with the auto encoder model and served to indicate how many anomalies each threshold level would classify.

Initially a set value was used as the threshold. The value of 1 was chosen as the starting value as that would indicate that the prediction was wrong by a whole unit of whichever sensor was being tracked at the time. This posed a problem with sensors that have a large variation. While a change of 1 in the temperature sensor is relatively significant, in sensors such as humidity this is not a significant change at all.

Below are the results of testing different solid value thresholds on the Bluetooth data sets.

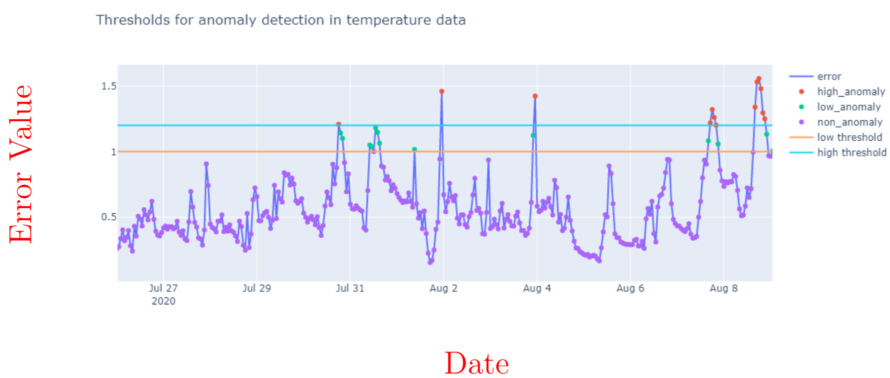


Figure 10: Anomalies detected in Property 1 Bluetooth temperature data using threshold with high of 1.2 and low of 1

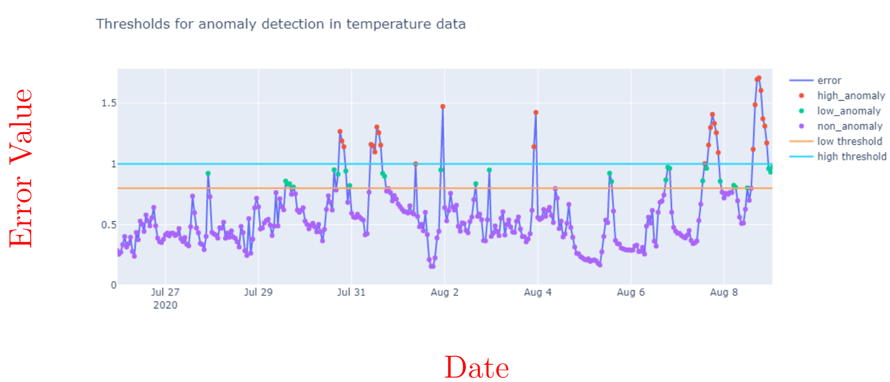


Figure 11: Anomalies detected in Property 1 Bluetooth temperature data using threshold with high of 1 and low of 0.8

Solid Value	Identified Anomalies
1.2	27
1	53
0.8	116

Table 24: Identified anomalies in Property 1 temperature data for different solid value thresholds

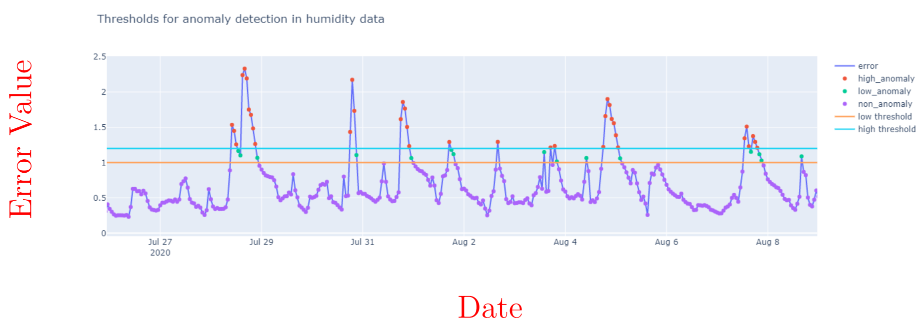


Figure 12: Anomalies detected in Property 1 humidity data using threshold with high of 1.2 and low of 1

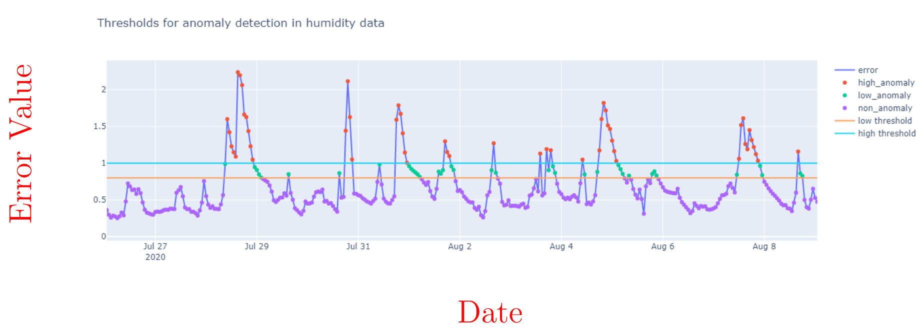


Figure 13: Anomalies detected in Property 1 humidity data using threshold with high of 1 and low of 0.8

Solid Value	Identified Anomalies
1.2	74
1	112
0.8	183

Table 25: Identified anomalies in Property 1 humidity for different solid value thresholds

After some consideration it was decided to change to percentage measure for the thresholds. Rather than a set value, a percentage of how much error is allowed was selected as the threshold. This allows for the same threshold to be used independent from a specific sensor, ensuring similar results for each of the different data types in the property.

The percentage measure works by calculating the mean absolute error rate of the model when used on the training data. This is then considered the general error rate for the model and is considered a baseline. Thresholds are selected by finding the value that a specific percentile of the errors in the training data are within, and then applying that value as a threshold when the system is used on the actual data.

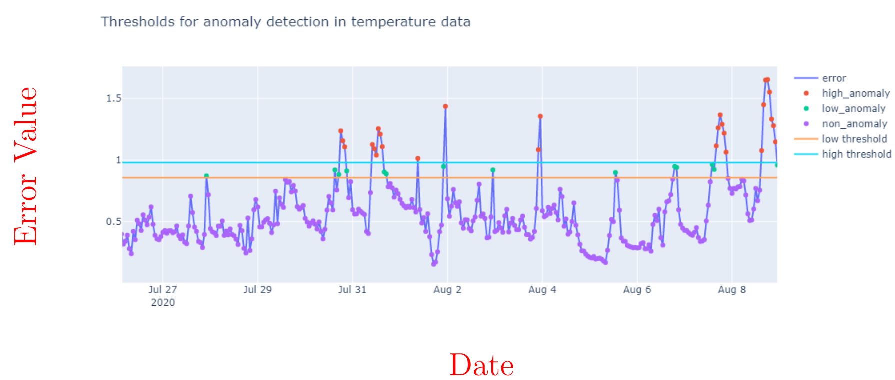


Figure 14: Anomalies detected in Property 1 temperature data using threshold with high of 5% and low of 10%

Percentage Value	Identified Anomalies
5%	57
10%	86
15%	117
25%	190

Table 26: Table showing identified anomalies in Property 1 temperature for different percentage value thresholds

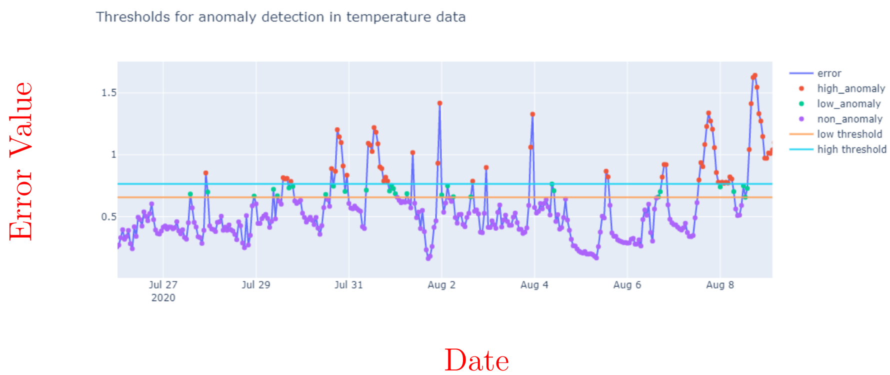


Figure 15: Anomalies detected in Property 1 temperature data using threshold with high of 15% and low of 25%

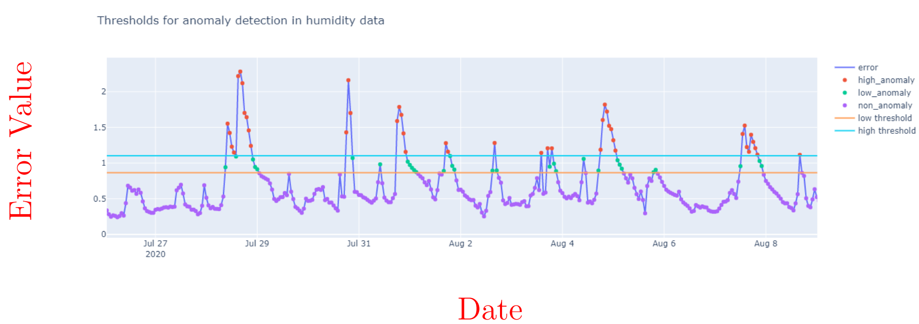


Figure 16: Anomalies detected in Property 1 humidity data using threshold with high of 5% and low of 10%

Percentage Value	Identified Anomalies
5%	98
10%	157
15%	214
25%	321

Table 27: Table showing identified anomalies in Property 1 humidity for different percentage value thresholds

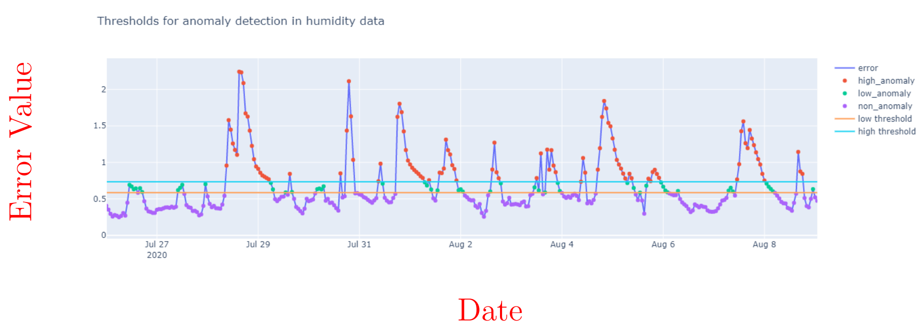


Figure 17: Anomalies detected in Property 1 humidity data using threshold with high of 15% and low of 25%

8.2.1 Threshold Conclusion

From the results above we can see that a percentage based system works best for creating anomaly thresholds. This allows for the system to adapt to different data types, without requiring for a user to manually adjust the system each time.

Using the testing results we decided on default values. It was decided for the current system to use 5% as a high threshold and 10% as a low threshold, as according to the Safehouse team these provided the most manageable number of anomalies.

The percentage of anomalies detected is able to be customised in the user interface, allowing for the system to be adapted to the users needs. This allows for the user to modify how strict the thresholds are, thus allowing the user to classify more or less points as anomalies. This may be useful in situations where an "at risk" person is being monitored, as what may be a minor anomaly for someone else, may indicate a severe problem for them.

8.3 Single Sensor Anomaly Detection

8.3.1 Details

The purpose of the system is to display the high and low anomalies detected in the properties. The number and location of which will give a good idea of how effectively the system is able to work on the different properties.

To provide a more objective view of the results the RMSE of using the system on each of the sensors will be provided. As suggested by other works (Chai and Draxler 2014), this will provide a more generalised view of how the system is performing in each instance. The RMSE can also be compared between the two properties to detect instances where the model performed more favourably in one property over the other.

Below are the test results of running the generic anomaly detection system on the different environmental sensor types in the two properties. For the sake of these tests the LSTM Autoencoder is used, as this was found to provide consistent results in a short time span.

8.3.2 Temperature

As can be seen, the system picked up on the largest peaks and drops in this sensor and identified them as anomalies. At key points, such as on August the 9th, the system was also able to detect where the temperature was beginning to ascend at an abnormal rate, resulting in a spike.

Of note is that the temperature itself did not change much during the time period, generally remaining between 20 and 24 degrees, with some peaks and drops out of this range. Despite this several anomalies were still detected in the 20 to 24 degree range. Generally this range would be considered safe for the average property, so while anomalies have been detected they may not require immediate attention.

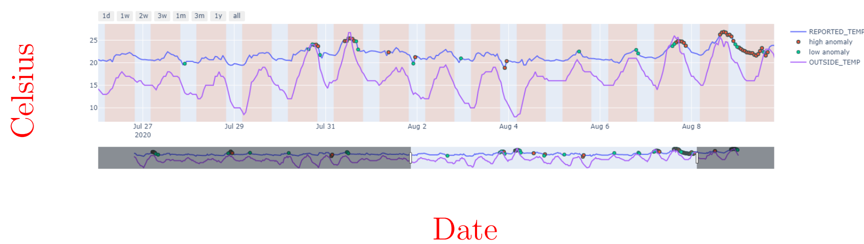


Figure 18: Regular anomaly detection in property 1 temperature

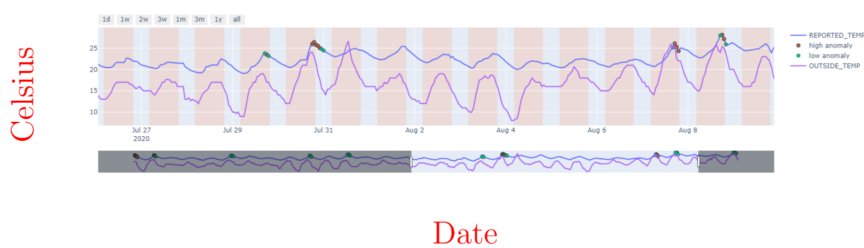


Figure 19: Regular anomaly detection in property 2 temperature

8.3.3 Humidity

The charts below also show that several of the highs and lows in the data were identified as expected. Dips that occurred during the night were regularly detected by the system, as well as sharp rises such as on July the 29$^{\text{th}}$.

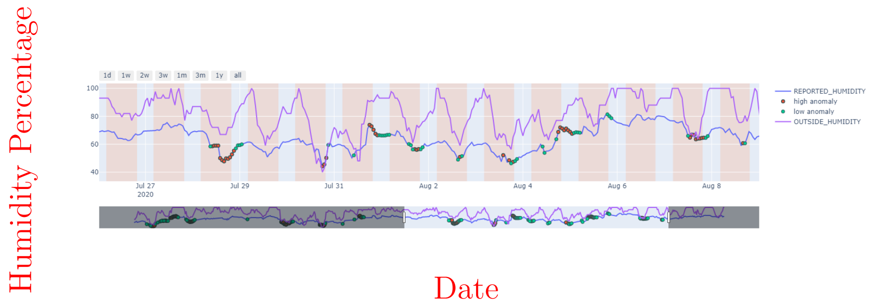


Figure 20: Regular anomaly detection in property 1 humidity

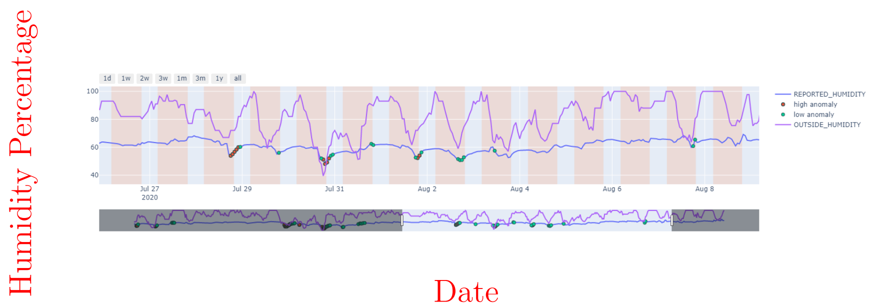


Figure 21: Regular anomaly detection in property 2 humidity

8.3.4 Light

The light sensors showed a similar pattern each day in property 1. Each of the peaks have been correctly identified as anomalies. This same pattern was observed in the property 2 with all sharp peaks being identified as anomalies, while smaller values were generally ignored.

Neither of the properties appear to have much light at night, resulting in no night anomalies being detected during this time period. The lack of anomalies due to a lack of light suggests that this same pattern was present in the data used for training the model.

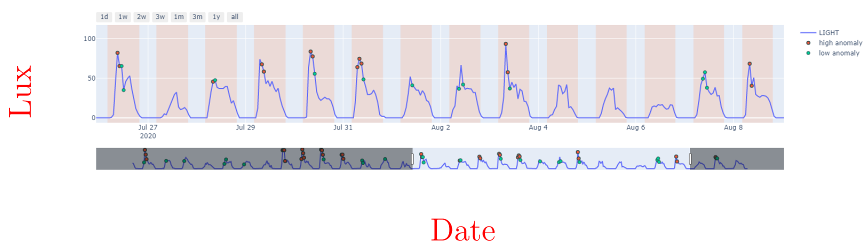


Figure 22: Regular anomaly detection in property 1 light

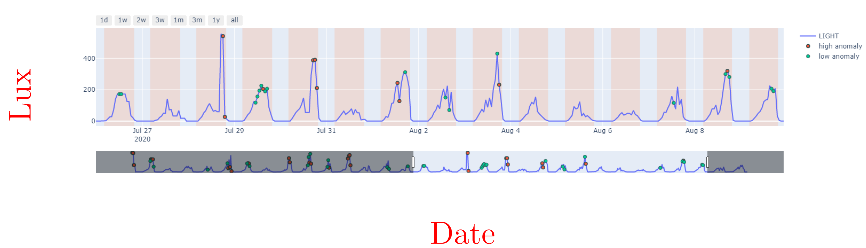


Figure 23: Regular anomaly detection in property 2 light

8.3.5 Motion

Unlike the other sensors that use the average for the hour, the motion sensor detects the total number of times that the sensor detected motion within the hour, giving a good idea of how much activity was performed in the room.

The majority of anomalies were detected when there was a small amount of activity, while points where there were high or low spikes did not register as anomalies. This can be explained as over time the system will have encountered a high number of lows and peaks in the data and will learn to register those points as normal.

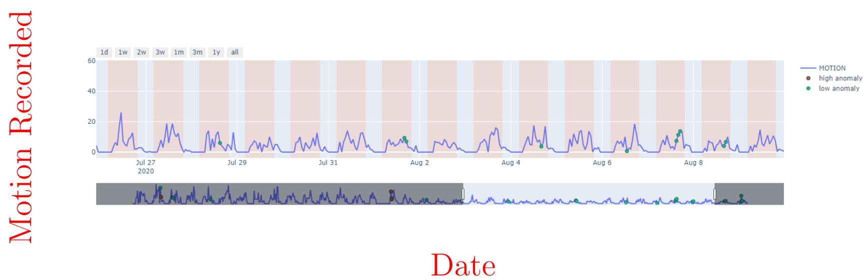


Figure 24: Regular anomaly detection in property 1 motion

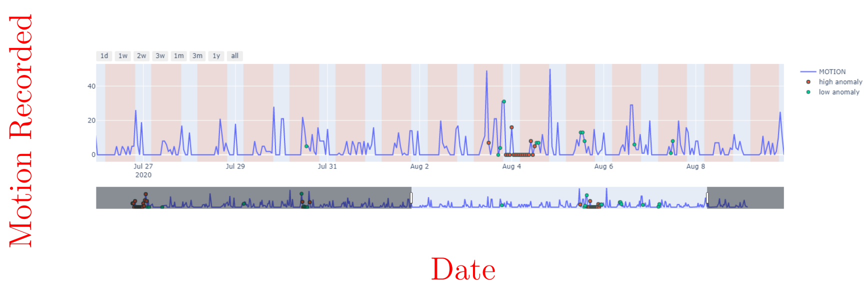


Figure 25: Regular anomaly detection in property 2 motion

8.4 Combined Anomaly Detection

8.4.1 Method

In order to allow for a multi sensor approach an LSTM Time Series prediction model is being used. This model works similarly to the LSTM system discussed in this paper till this point, and will use the same parameters. It takes an input and attempts to predict future readings for that input based on what has been given to it.

The combined sensor approach uses a multivariate input in order to predict the future readings for a single primary input. The model takes in a single primary input, this will be the sensor which the model will attempt to predict future readings for. Several secondary inputs are also added to the model, the readings of these sensors will be used in order to attempt to predict the primary inputs future readings.

As the inputs are all scaled prior to being used by the model, they are all equally weighed in the system. This stops excessively high readings in one sensor from negatively affecting the result of the model. This also allows for sensors that have vastly different reading values to be used together.

The accuracy of this model is dependent on how closely linked the results of the sensors used as inputs are, as past correlation is being used to detect the future. As can be seen further on, obvious examples such as temperature and humidity are very closely linked and can therefore be used in order to make accurate predictions.

As with the earlier model, anomalies are detected based on how different the actual future values are in comparison to what was predicted by the model. Similar thresholds are also in place in order to evaluate anomalies as either severe or minor.

An extra benefit of this model is that we're able to see whether the anomalies detected were anomalies because their results were higher or lower than expected. This is useful as it gives insight as to which way things are changing and makes it easier to work out what the issue is. For example a reading that's lower than expected in the temperature data may indicate a sharp drop, while a higher temperature than expected may indicate a sharp rise.

8.4.2 Temperature with Humidity

As we can see from the RMSE and R2 score this algorithm was able to effectively predict the future temperature with a great degree of accuracy.

Property	MAE	RMSE	R2
Property 1	0.217	0.301	0.928
Property 2	0.141	0.232	0.906

Table 28: Results of utilising a combined approach of temperature with humidity

An R2 score so close to 1 suggests that the predictions made were very similar to the actual reading for the model. In general a reading above 0.7 for R2 is considered very accurate. This accuracy is also shown by the RMSE being as low as it is. Similar results can be observed in the second property. The results of running this model on that property were remarkably similar to the results of the first.

As shown in the charts, the system was able to successfully identify a large amount of anomalies in the first property. Of note however is that a couple of major rises were missed by the system. Namely the rise on the 2^{nd} of August and at the end of the chart on the 13^{th}.

In this aspect the system appears to have done a better job of predicting anomalies in the second property. Managing to detect all of the highest peaks as anomalies, however it did miss a couple of the earlier dips.

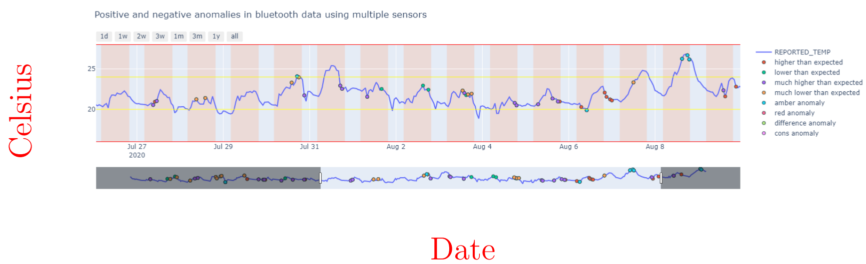


Figure 26: Anomaly detection in property 1 temperature supplemented by humidity

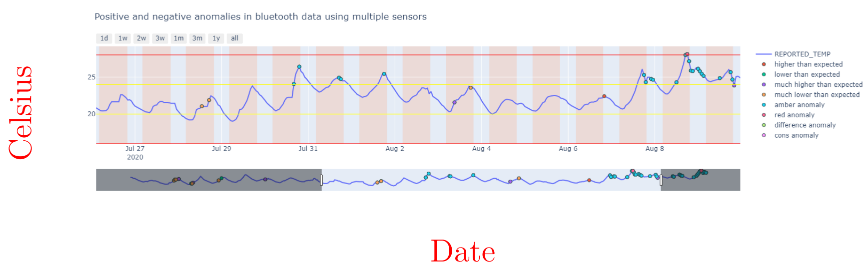


Figure 27: Anomaly detection in property 2 temperature supplemented by humidity

When looking at the total number of anomalies in both properties we can see that the number detected in both was nearly identical.

The notable difference between the two properties is that a near equal number of positive and negative anomalies were found in property 1, while property 2 had significantly more positive than negative anomalies. This may explain why some of the dips were not identified in the second property chart.

Type	Occurrences
Total Data	607
Total Anomalies	60
Positive Anomalies	31
Negative Anomalies	29
High Positive Anomalies	22
High Negative Anomalies	19

Table 29: All readings in property 1 temperature supplemented with humidity

Type	Occurrences
Total Data	607
Total Anomalies	58
Positive Anomalies	45
Negative Anomalies	13
High Positive Anomalies	36
High Negative Anomalies	9

Table 30: All readings in property 2 temperature supplemented with humidity

The RMSE when predicting using multiple sensors was significantly lower than in the single sensor prediction. The RMSE of using the single sensor system on temperature was 0.657, in comparison to the 0.295 when using temperature combined with humidity, giving it an RMSE of less than half of that when using a single sensor to make predictions. This indicates that the multi sensor prediction was significantly more accurate than when only a single sensor was used.

The lower RMSE and R2 so close to 1 also suggest that the temperature and humidity in the property are very closely linked and can therefore be used together to more accurately predict future results.

Combining the different types of anomalies gives a total of 60 anomalies when using the combination of temperature and humidity. In comparison when detecting anomalies on temperature alone, 81 anomalies were detected. This increase is likely due to the higher RMSE in the single sensor prediction, and indicates that some of those detected anomalies are potential false positives.

In conclusion it would appear that a combination of temperature and humidity provides more accurate results than when using a single sensor. This comes with drawbacks however. The most notable of these drawbacks is that it takes significantly longer to use the system on multiple sensor types. This is a particularly big problem when we expect to run the system regularly so as to be able to detect anomalies as they happen. Another drawback is that a lot more data is required. The data of two relevant sensors is needed in order to detect these anomalies, whereas the single sensor system only needs one.

8.4.3 Temperature with Light

Combining the temperature sensor data with the light sensor data provides significantly different results. As can be seen in the tables below, this combination of sensors provides worse results than using the individual temperature sensors in both properties. This is because of the lack of correlation between the two sensors, as temperature and light in the room are not reliant on each other, unlike temperature and humidity which are.

Property	MAE	RMSE	R2
Property 1	0.278	0.528	0.657
Property 2	0.238	0.536	0.644

Table 31: Results of utilising a combined approach of temperature with light

Type	Occurrences
Total Data	607
Total Anomalies	53
Positive Anomalies	23
Negative Anomalies	30
High Positive Anomalies	22
High Negative Anomalies	24

Table 32: All readings in property 1 temperature supplemented with light

Type	Occurrences
Total Data	607
Total Anomalies	48
Positive Anomalies	30
Negative Anomalies	18
High Positive Anomalies	26
High Negative Anomalies	12

Table 33: All readings in property 2 temperature supplemented with light

8.4.4 Temperature with Sound

As with the combination of temperature and light, this combination was less effective than the individual temperature sensors in the two properties. This is likely due to the same reason as the correlation between the temperature and sound readings are unlikely to be tightly linked.

Property	MAE	RMSE	R2
Property 1	0.397	0.577	0.588
Property 2	0.560	0.734	0.290

Table 34: Results of utilising a combined approach of temperature with sound

Type	Occurrences
Total Data	607
Total Anomalies	73
Positive Anomalies	28
Negative Anomalies	45
High Positive Anomalies	20
High Negative Anomalies	35

Table 35: All readings in property 1 temperature supplemented with sound

Type	Occurrences
Total Data	607
Total Anomalies	48
Positive Anomalies	33
Negative Anomalies	19
High Positive Anomalies	26
High Negative Anomalies	12

Table 36: All readings in property 2 temperature supplemented with sound

8.4.5 Humidity with Temperature

Using the temperature sensor data to supplement the humidity sensor provides similar results to when the sensors are combined in the opposite way, that is temperature supplemented by humidity. In this way the combination provides similar results to when the humidity sensor is used by itself. This is likely due to the obvious connection between humidity and temperature in a property.

Property	MAE	RMSE	R2
Property 1	0.207	0.275	0.890
Property 2	0.098	0.138	0.968

Table 37: Results of utilising a combined approach of humidity with temperature

Type	Occurrences
Total Data	607
Total Anomalies	68
Positive Anomalies	42
Negative Anomalies	26
High Positive Anomalies	29
High Negative Anomalies	20

Table 38: All readings in property 1 humidity supplemented with temperature

Type	Occurrences
Total Data	607
Total Anomalies	65
Positive Anomalies	46
Negative Anomalies	19
High Positive Anomalies	30
High Negative Anomalies	13

Table 39: All readings in property 2 humidity supplemented with temperature

8.4.6 Humidity with Light

As with the temperature, combining humidity with the light sensor also provided worse results than using the humidity sensors of both properties individually. The results from the two properties were relatively similar, indicating no clear differences between the two when it came to this combination of sensors.

Property	MAE	RMSE	R2
Property 1	0.248	0.442	0.760
Property 2	0.243	0.532	0.650

Table 40: Results of utilising a combined approach of humidity with light

Type	Occurrences
Total Data	607
Total Anomalies	59
Positive Anomalies	22
Negative Anomalies	37
High Positive Anomalies	19
High Negative Anomalies	29

Table 41: All readings in property 1 humidity supplemented with light

Type	Occurrences
Total Data	607
Total Anomalies	49
Positive Anomalies	29
Negative Anomalies	20
High Positive Anomalies	26
High Negative Anomalies	12

Table 42: All readings in property 2 humidity supplemented with light

8.4.7 Humidity with Sound

Similarly to when the humidity was combined with light, the results of combining humidity with light also provided poor results. Notably however was the greatly reduced accuracy of the model in the second property. While there are likely multiple reasons for this difference, the largest is that the noise in the property does not appear to match any other sensor data. This may be because of the location of the sensor in the room, or the presence of external noise that is not also in the first property.

Property	MAE	RMSE	R2
Property 1	0.414	0.595	0.562
Property 2	0.593	0.784	0.190

Table 43: Results of utilising a combined approach of humidity with sound

Type	Occurrences
Total Data	607
Total Anomalies	71
Positive Anomalies	37
Negative Anomalies	34
High Positive Anomalies	22
High Negative Anomalies	27

Table 44: All readings in property 1 humidity supplemented with sound

Type	Occurrences
Total Data	607
Total Anomalies	69
Positive Anomalies	45
Negative Anomalies	24
High Positive Anomalies	34
High Negative Anomalies	16

Table 45: All readings in property 2 humidity supplemented with sound

8.5 Conclusion

Initially the evaluation of whether a point is considered an anomaly is the same across the different sensor types. Allowing the system to work on most sensor types provided, but failing to account for the context of the different sensor types. This could be a future problem as the majority of values detected remained within what is generally considered a safe range.

The system performed best on the temperature and humidity sensors. Each of these sensors provide continuous data that changes gradually over time, while keeping to some pattern based on the time of day. For these reasons both sensors are prime candidates for machine learning that aims to find patterns, and points outside of those patterns.

In comparison the system performed poorly on the light and motion sensors. Likely due to the irregular changes in their value as the results are entirely dependent on human interaction which can be erratic.

From the tests we performed a combination of sensors provided better results than a single sensor solution in cases where the sensors used were relevant in some way. This can be most clearly seen in the temperature and humidity combinations where the results were more accurate than single sensor solutions for those sensors. However the drawbacks of this method were severe, requiring a significant amount of data and time in order to obtain these results. Using sensors that are not as deeply related also reduced the quality of the results, with the combination of sensors performing worse than the individual sensors. As seen in the instance where the temperature sensor data was supplemented by the light or sound sensor data.

9 Anomaly Detection With Additional Parameters

9.1 Additional Conditions

The primary improvement we aimed to implement was a range of extra conditions that were provided by the Safehouse team.

A key piece of feedback we received was that the system was too sensitive, and that this was resulting in an unrealistic number of points being identified as anomalies. This is problematic as the workload for the person monitoring the occupant would be very high as they would need to manually check each of the detected anomalies. Reducing this to a manageable number was a task that the Safehouse team defined as a top priority.

Safehouse provided a number of extra conditions to help narrow down the number of anomalies, and to help identify which anomalies are the most important. As well as these extra conditions, several suggestions have been provided for specific things that should be looked for in the property that indicate an anomaly or danger to the occupant. These may detect unique anomalies that the system was unable to detect by itself, or enhance the knowledge we have on the existing anomalies.

9.1.1 Traffic Light Range

In order to narrow down the number of anomalies detected a traffic light system was implemented. This categorises points within three different zones based on how comfortable they would be for the occupant and the property as a whole.

The traffic light system was provided by the experts from the Safehouse team. The levels used for this system were defined based on their own research and expertise in the subject. As the traffic light system is different for each of the sensor types, Safehouse provided traffic light charts for several different sensors in the properties. These denote the ranges for the different danger levels, as well as the common problems that may occur if they are exceeded.

An obvious limitation of this system is that it takes a "one size fits all" approach to classifying data. While the traffic light system provides general guidelines for comfortable and safe levels of particular readings in a property, what is comfortable for one person, may not be for another. For this reason the Safehouse team have worked to ensure that the guidelines provided are suitable for the majority of people.

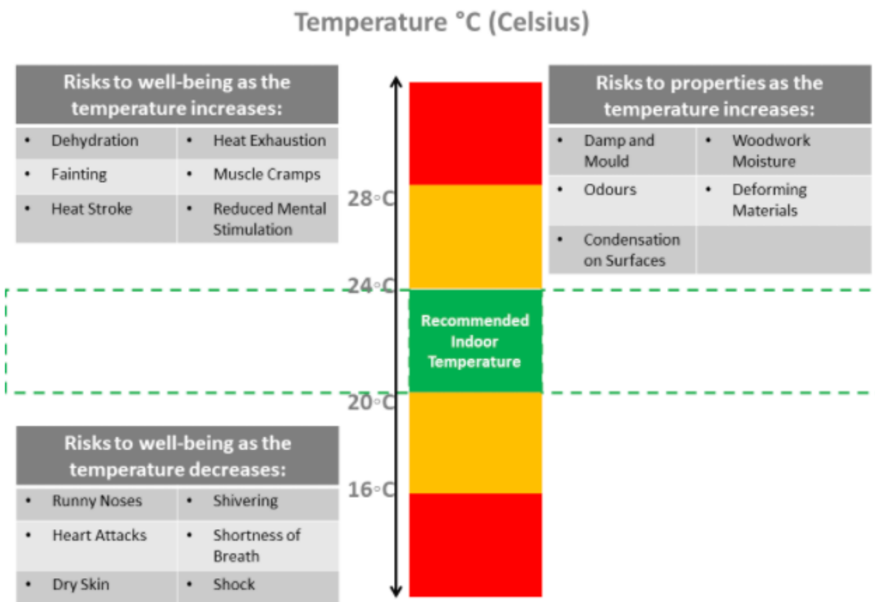


Figure 28: Traffic light range for temperature

Points within the green zone are considered normal readings and can be safely ignored. These are points that are within a comfortable temperature range for the occupant and should have no adverse effects on the occupant or property.

Points within the amber zone are detrimental, but should not have any life risking impacts on the occupant. Temperatures within this range may result in the occupant becoming uncomfortable, as well as some minor risks to the property.

Points within the red zone are considered major problems and should be avoided whenever possible. Allowing the temperature to remain within this range can result in severe impacts to the well being of the occupant, as well as long lasting damages to the property.

As with temperature, humidity also has a comfortable range that the property needs to remain within. Any values outside of this range could reveal, or cause, serious problems in the property. The range for the different thresholds and the problems they may cause are illustrated in the diagram.

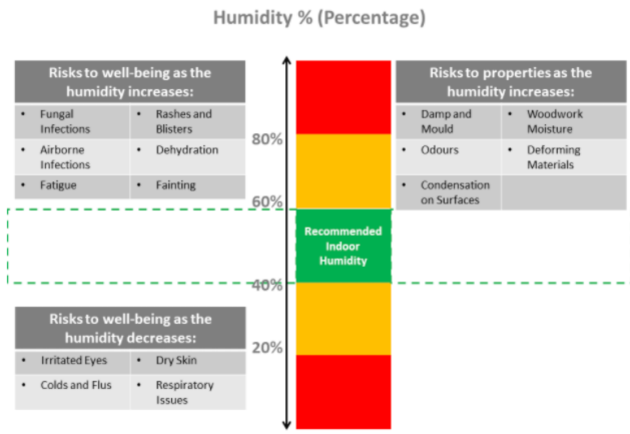


Figure 29: Traffic light range for humidity

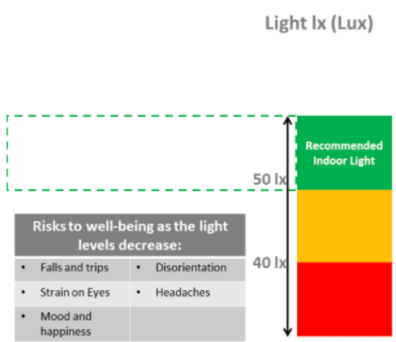


Figure 30: Traffic light range for light

Light is another value that has an ideal range. Unlike temperature and humidity this is a level that the sensor should remain above. It's important to maintain a good light level when the occupant is in the room, particularly at night so as to avoid any potential accidents, but also during the day to reduce disorientation and headaches.

While less extreme of an issue, low light levels in the property can also have a negative effect on the occupants overall mood, diminishing their quality of life.

9.1.2 Constant

While using the traffic light system in conjunction with the basic anomalies will allow us to more accurately gauge the severity of anomalies, the traffic light system can also allow for other anomaly types to be detected. One such anomaly is when the value of a sensor remains within a dangerous range for a prolonged period of time.

This time period will be dependent on which sensor is being monitored. The specific time periods for each of the sensors were provided by the Safehouse team.

In the case of temperature the system was looking to detect instances where the temperature was above or below the red line for more than 24 hours without equalising. This is useful as the longer the temperature remains in the red zone the more likely that the occupant, or property, will suffer the possible negative effects from it.

For humidity the system is aiming to detect where it has remained in the red zone for a period of 24 hours or longer. It's important to detect when this occurs as high levels of humidity for a prolonged period of time can cause mold growth within the house, which may then result in future health concerns for the occupant.

Consecutively remaining in the red zone for the light sensor indicates a different kind of anomaly than when this is detected in the other sensors. Remaining in the red zone for over 24 hours indicates that the lighting in the room has not hit a level comfortable for the occupant. This can indicate one of a few things. It may indicate that the occupant has not entered the room for a long length of time. Which is problematic if the room is a kitchen or bedroom, which they would be expected to use regularly. It may also indicate that the lights are not working, or that the occupant is simply failing to turn the lights on.

9.1.3 Difference in Values

Extra conditions were suggested in order to detect sharp drops, or rises, that may have been missed by the main system. The difference between two corresponding values must not exceed a set threshold value that indicates a drastic change in what was detected. This threshold will need to be adapted for each of the different sensor types that are being used as a different drop, or rise, is appropriate for different sensor types.

In the temperature sensor we are looking for a change of 4 degrees or more in an hour. This would be a severe change and would indicate that something is seriously wrong within the property.

9.1.4 Percentage Change

One of the conditions the Safehouse team asked to be implemented was to check when the percentage change between the inside and outside sensors differ by 20% or more. This is particularly important in the temperature and humidity sensors as it shows a strange change in either the inside, or outside, atmosphere.

In general the occupant should be attempting to keep the room in a comfortable state at all times. To do this they will need to adjust the temperature to match the changes outside. This extra condition will check whether this is the case.

9.1.5 Outside Comparison

The system checks for instances where the outdoor humidity drops below the indoor humidity. This is a problem as it may indicate a problem in the property, particularly related to the heating and insulation.

9.1.6 Motion in the Dark

An extra condition exclusive to the motion sensor is the need to check for instances where the occupant is moving in the dark. This is dangerous for a number of reasons, for example the occupant is more likely to fall if they're unable to see where they're going. The occupant failing to turn on the light may also signify deeper underlying problems as this would definitely fall out of the range of usual activity.

To detect this kind of anomaly three different things will be monitored. These are whether the occupant is moving, whether it's night time, and whether the lighting within the room is sufficient. If all three conditions are met then it indicates that the occupant is likely moving around the house in the dark.

9.2 Temperature Results

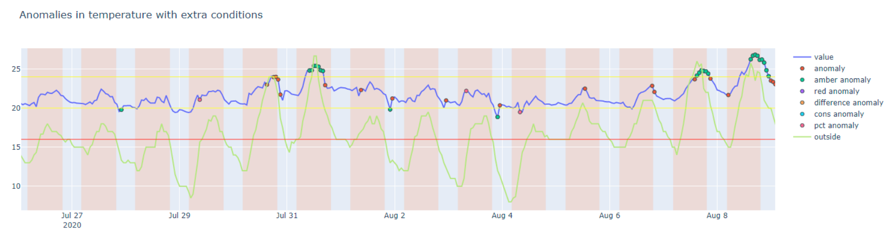


Figure 31: Property 1 Temperature with all conditions

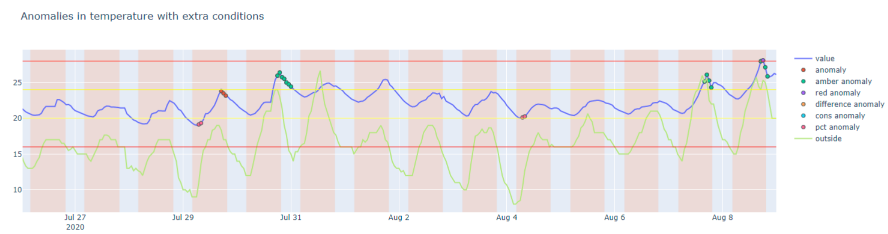


Figure 32: Property 2 Temperature with all conditions

As shown in the previous results, the system does a good job of detecting a large number of anomalies, particularly in areas where a large peak, or drop, has occurred in the data set.

One of the issues with the existing method of detecting anomalies is that the context related to each of the sensor types is ignored. While this allows the model to work on several different sensors with ease, it does struggle with cases that while unusual may be acceptable because of the type of readings being monitored. An example of this is found in the temperature, where even in the event of a sharp drop, so long as the temperature remains within a comfortable range this is perfectly acceptable and as such does not need to be detected as an anomaly.

9.2.1 Standard anomalies

Points are detected as anomalies when they are above the low threshold of error as defined earlier on. Points detected as low and high anomalies were combined together in order to avoid extra confusion that may occur by having the extra parameters as well as low and high anomalies.

As shown in charts a large number of points were identified as anomalies in the data set, particularly towards the end of the data such as on August the 9^{th}, where there is a large peak followed by a sharp drop.

9.2.2 Amber and red anomalies

In order to narrow down the number of anomalies detected a traffic light system was implemented. This categorises points within three different zones based on how comfortable they would be for the occupant.

The traffic light system has been implemented into the system. The boundaries for each of the sections are indicated by the yellow and red lines on the chart. Anything within the two yellow lines is green, anything between the yellow and red is amber and anything outside of the red lines are red.

To make this clearer anomalies that fall within the amber or red ranges have been labeled as such, while potential anomalies that fall within the green zone have been removed altogether.

This has two notable effects:

- Red anomalies can be easily identified as more severe problems that need to be resolved.

- False positives are reduced, as anomalies detected within the green zone have been removed, thus reducing the number of points that the carer needs to be aware of.

One point of note is that the sharp drop detected on August 9^{th} remains within the ideal conditions for the house, as set out by the traffic light system provided by the Safehouse team.

9.2.3 Difference anomalies

In this instance we're looking for a change of 4 degrees within the space of an hour. However as can be seen in the chart below no instances of this happening have been found.

While it's possible that there were no suspicious drops in the data, future adjustments may be needed to accurately detect sharp drops. Checking for changes over the course of 2 hours rather than just 1 may provide more insight.

9.2.4 Constant anomalies

While no instances have been detected in either of the test properties, extra conditions have been put in place in order to check when the temperature has stayed above or below the red comfort threshold for a significant amount of time.

9.2.5 Percentage anomalies

If the difference in the rate of change between the indoor and outdoor sensors is more than 20% then an anomaly is flagged. Both properties had several instances of this as shown by the identified anomalies.

This is useful for identifying instances where the temperature within the house is not changing consistently with outside, potentially indicating an unusual drop or rise in temperature.

9.2.6 Overall

Type	Occurrences
Total Data	751
Low Anomaly	35
High Anomaly	46
Anomaly	81
Amber Anomaly	36
Red Anomaly	0
Difference Anomaly	0
Consecutive Anomaly	0
Difference Percentage Anomaly	8

Table 46: Table showing all readings in property 1 temperature

These are the number of anomalies of different types found within both properties living room temperature data.

Type	Occurrences
Total Data	751
Low Anomaly	42
High Anomaly	26
Anomaly	68
Amber Anomaly	33
Red Anomaly	2
Difference Anomaly	0
Consecutive Anomaly	0
Difference Percentage Anomaly	10

Table 47: Table showing all readings in property 2 temperature

Of the 751 points of data in the first property, 35 were identified as low anomalies and 46 were identified as high anomalies, giving a total of 81 anomalies in the data.

Of the 81 detected anomalies 36 were above the amber threshold, and none were above the red. The fact that none of the anomalies were above the red threshold indicates that the occupant is probably not at a severe risk and may not need immediate attention. Depending on how the amber anomalies are grouped these may warrant further investigation as they indicate conditions that may result in some discomfort.

Similar results were found in the second property. The second property had fewer overall anomalies, but had two instances where red anomalies were detected. These are higher priority anomalies and indicate events that should be looked into.

No sharp drops over a one hour period were detected in the data as indicated by the lack of difference anomalies.

There were no consecutive anomalies in either of the properties, therefore the temperature in the homes was not kept at a dangerous level for a significant period of time.

A similar number of instances were detected in both properties where the indoor and outdoor temperature were not changing at a similar rate. As the number is so similar these may indicate a large drop, or rise, in the outside temperature rather than anything changing inside the house. Nevertheless it may be worth looking into these anomalies to ensure that the occupant is changing the temperature in their home to accommodate for outside.

9.2.7 Specialised compared to standard

Type	Occurrences
Total Data	751
Difference Anomaly Low	0
Difference Anomaly High	0
Consistent Anomaly Low	0
Consistent Anomaly High	0
Difference Percentage Low	0
Difference Percentage High	0

Table 48: Table showing matching readings in property 1 temperature

Type	Occurrences
Total Data	751
Difference Anomaly Low	0
Difference Anomaly High	0
Consistent Anomaly Low	0
Consistent Anomaly High	0
Difference Percentage Low	0
Difference Percentage High	0

Table 49: Table showing matching readings in property 2 temperature

As can be seen from the results there were no instances where the original system was able to detect the unique extra anomalies suggested by the Safehouse team in either of the properties.

This is inconclusive for this sensor type however as two of the anomaly types, difference and consecutive, were not detected in the data and the number of different percentage anomalies was low.

9.3 Humidity Results

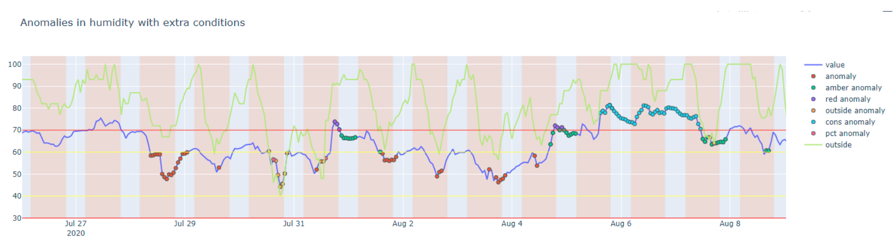


Figure 33: Property 1 Humidity with all conditions

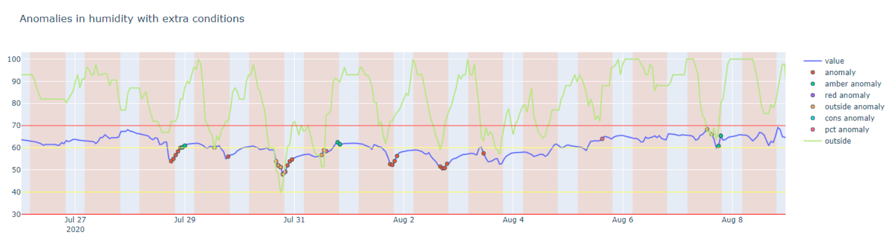


Figure 34: Property 2 Humidity with all conditions

As with the temperature sensor, there are a several humidity conditions that need to be checked in order to ensure that there are no problems in the house. These vary slightly from the conditions that have been set for the temperature sensor.

9.3.1 Amber and red anomalies

Unlike with temperature, several of the humidity anomalies crossed the red threshold, indicating anomalies that could have a significant impact on the occupant. These would be the most important points to identify as they may require immediate attention.

9.3.2 Outside anomalies

One of the issues exclusive to humidity is that the outside humidity should not be below the indoor humidity.

The system now checks for when the outside levels drop below those of inside. There are a few instances in the monitored period where this occurs.

9.3.3 Constant anomalies

Unlike with temperature, several points in the humidity data were detected after remaining above the red threshold for 24 hours. These indicate prolonged problems in the property that could have long term effects on the occupant and their quality of life.

9.3.4 Percentage anomalies

As with the temperature sensor, it's important to be able to tell when the rate of change in humidity crosses a certain threshold, as this would indicate a big change in the conditions of the indoor sensors in comparison to outside.

9.3.5 Overall

A significantly higher number of anomalies were detected in the humidity sensor over temperature, particularly in the first property. Each of the different anomaly types that the system was looking for were detected in the first property, with the second only missing red and consecutive anomalies. This indicates one of two things. Either the system is too sensitive for the humidity sensor, or the household has had serious issues with humidity.

One hundred and fifty five total anomalies were detected in the data. Without applying the extra conditions this provides a lot of points that would need to be examined to check their severity. The extra conditions have

Type	Occurrences
Total Data	751
Anomaly	155
Amber Anomaly	48
Red Anomaly	13
Difference Anomaly	39
Outside anomaly	23
Consecutive Anomaly	45
Difference Percentage Anomaly	8

Table 50: Table showing all readings in property 1 humidity

Type	Occurrences
Total Data	751
Anomaly	68
Amber Anomaly	10
Red Anomaly	0
Difference Anomaly	10
Outside Anomaly	31
Consecutive Anomaly	0
Difference Percentage Anomaly	9

Table 51: Table showing all readings in property 2 humidity

helped to cut this down to a more manageable number, while also potentially highlighting why they were identified as anomalies in the first place.

Slightly less than a third of the anomalies detected were within the amber zone. This drastically reduces the number of anomalies that would need to be checked, while ensuring that potentially important anomalies are marked.

Unlike temperature several anomalies were in the red zone in the first property, with 13 in total. These are the key anomalies that the person monitoring would want to be aware of. Not only do these indicate that an anomaly has occurred, but also that the environment within the house has reached dangerous levels.

A large number of points in the data were identified as difference anomalies. These are sharp rises or drops in the humidity value within the space of an hour. This is a problem because it indicates that the humidity within the house is fluctuating at unnatural levels.

One of the unique conditions of humidity that needs to be monitored is to ensure that the outside levels are higher than inside. As can be seen from the table there were 23 times where this was not the case.

45 consecutive anomalies were detected in the first property. This indicates that the house was consistently kept at an uncomfortable level. Reacting before this number of consecutive anomalies is detected will not only help keep the occupant healthy and happy, but may also prevent permanent damage to the house. Interestingly the second property experienced none of this type of anomaly.

The differential percentage anomalies indicate there were 8 and 9 instances, in the first and second properties respectively, where the difference in change of the indoor and outdoor humidity varied greatly. Again this is very similar to the results of this type of anomaly in the temperature for the properties, further indicating that it's more likely linked to outside of the house changing rather than inside. The fact that the two properties are so similar also indicates that this is not an unusual change.

9.3.6 Specialised compared to standard

Type	Occurrences
Total Data	751
Difference Anomaly Low	12
Difference Anomaly High	6
Outside Anomaly Low	9
Outside Anomaly High	9
Consistent Anomaly Low	2
Consistent Anomaly High	0
Difference Percentage Low	2
Difference Percentage High	2

Table 52: Table showing matching readings in property 1 humidity

Type	Occurrences
Total Data	751
Difference Anomaly Low	3
Difference Anomaly High	2
Outside Anomaly Low	7
Outside Anomaly High	5
Consistent Anomaly Low	0
Consistent Anomaly High	0
Difference Percentage Low	4
Difference Percentage High	4

Table 53: Table showing matching readings in property 2 humidity

Unlike in the results from the temperature sensor there were several instances where the extra anomalies were detected by the base system. This is likely because of the sheer number of extra anomalies detected in the humidity data.

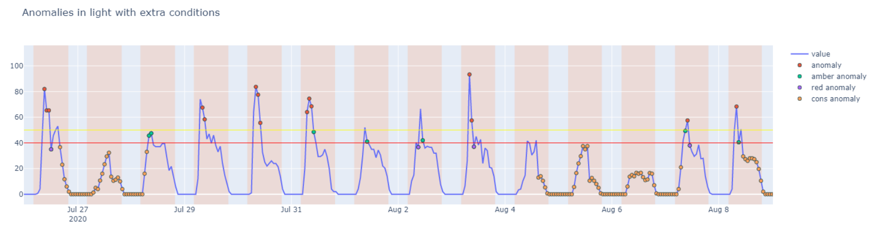
Interestingly despite there being a large variance in the extra anomalies found between the two properties, the number of extra anomalies that correlate with the standard anomalies was similar for the two.

Of note however is that very few of the consecutive anomalies were picked up on. This makes sense for the main system as it's looking for a divergence from a standard pattern. In cases where the value remains in a similar area for several consecutive hours then the original system would begin to recognise that pattern and not treat any of those points as anomalies.

9.4 Light Results

There were two extra conditions to be detected in this property, the traffic light system and the consecutive anomalies.

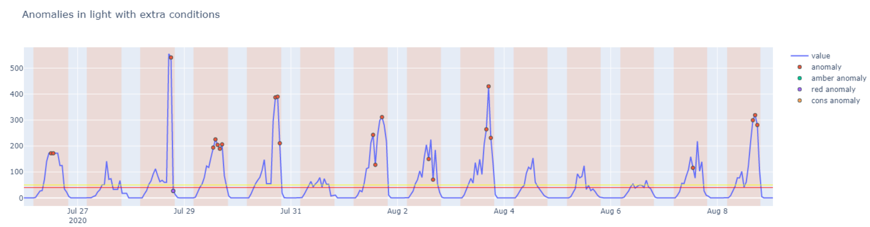
Lux



Date

Figure 35: Property 1 Light with all conditions

Lux



Date

Figure 36: Property 2 Light with all conditions

9.4.1 Overall

Type	Occurrences
Total Data	751
Low Anomaly	24
High Anomaly	30
Amber Anomaly	10
Red Anomaly	7
Consecutive Anomaly	194

Table 54: Table showing all readings in property 1 light

Type	Occurrences
Total Data	751
Low Anomaly	28
High Anomaly	31
Amber Anomaly	0
Red Anomaly	2
Consecutive Anomaly	0

Table 55: Table showing all readings in property 2 light

Both properties had instances of red anomalies, however only the first had any amber anomalies. In comparison to the standard anomalies, the number of amber and red anomalies were fairly low.

As amber and red anomalies are based on when light level is below a specific range, the lack of specialised anomalies detected is likely due to the light in the property being significantly higher than expected. Extra tuning may be needed here to check what the light levels should be at different points in the day.

9.5 Motion Results

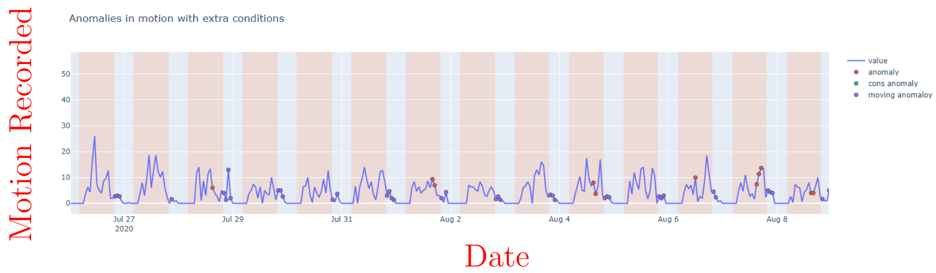


Figure 37: Property 1 Motion with all conditions

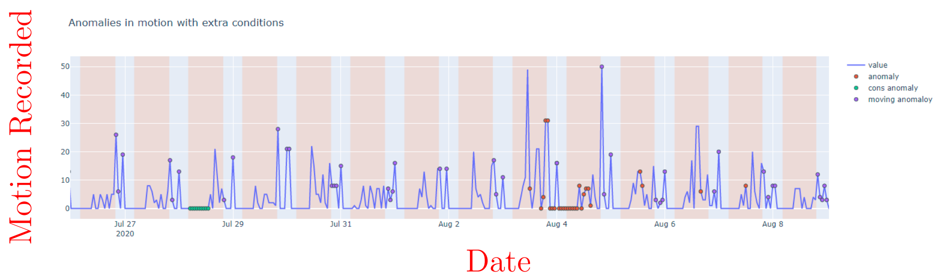


Figure 38: Property 2 Motion with all conditions

There are two extra details to look for in the motion sensors. The first is when the occupant performs no activity for a prolonged period of time. This may indicate that they're not able to move around for some reason and may require assistance. After discussion with the Safehouse team we chose to detect when the occupant had not moved for 6 hours or more in a row.

One of the main limitations with the motion sensor is that it's not available in every room. In the first property there are motion sensors in the bedroom and kitchen, while the second property only has a motion sensor in the kitchen. This allows us to pick up on activity in those rooms, but will fail to detect any activity happening in the rest of the house.

Unsurprisingly most days picked up 6 hours with no motion activity. This is mostly because of the night time where the occupant would be sleeping.

As can be seen from the chart there was only a single day in the first property where a long period with no movement was detected. We can also see that the lack of movement was early in the day which may suggest that the occupant had a late start, or perhaps the sunrise was earlier than usual.

Looking at the general motion after the anomalies we can see that the occupants' behaviour went back to normal after the morning. Suggesting that no further action would be needed at that time.

The second condition that the system is aiming to detect is more complex. This is when the occupant is moving around in the dark. As specified earlier, this type of anomaly is detected when the user is moving at night in a room with a low light level.

As can be seen from the results below there were a good number of times where this was the case in both properties. Particularly early in the night there were instances almost every day in both properties where this type of anomaly was detected.

Further refinement may be needed to set a later time frame before this type of anomaly is triggered. With more data we may also be able to see the difference in the winter where it becomes darker earlier and therefore would be more important that the property has sufficient light levels.

9.5.1 Overall

Type	Occurrences
Total Data	886
Low Anomaly	26
High Anomaly	2
Consecutive Anomaly	9
Moving Anomaly	104

Table 56: Table showing all readings in property 1 motion

Type	Occurrences
Total Data	747
Low Anomaly	60
High Anomaly	37
Consecutive Anomaly	23
Moving Anomaly	90

Table 57: Table showing all readings in property 2 motion

The lack of anomalies detected by the system in the first property suggests that the occupants' movement in the kitchen is similar each day. The second property however had vastly different results, with more than double the number of anomalies detected, suggesting the opposite.

There were 9 consecutive anomalies in the first property, signifying a prolonged period without movement in a single day. As all of these anomalies were grouped together, we can see that this was the only day in the dataset where the occupant did not move through the kitchen. Again the second property was vastly different, with consecutive anomalies on several days. In a realistic situation the number and frequency of these anomalies should result in some investigation by whoever is monitoring the property, as they may indicate that the occupant has not left their room for some reason.

9.6 Conclusion

Adding extra conditions significantly reduced the number of points that were detected by the system, in turn resulting in fewer false positives as identified by our industry experts. The problem with this however is that potential true positives may be missed in situations where the criteria is too strict. Further refinement of the criteria parameters may be needed to ensure that this is not the case going forward.

The most effective conditions were the traffic light system that was implemented. This allowed the system to automatically classify the expected severity of any anomalies detected based on the actual sensor values.

More specific conditions were demonstrably less effective, likely due to the fact that they were looking for indicators that would result in a severe problem, such as the temperature remaining too hot or cold for a large period of time. These conditions picked up on very few instances where they were triggered over the course of this study, and thus more testing would be needed to decide whether they are useful for the system going forward.

One instance where a specific anomaly was effective however was for checking motion in a specific period of time. This allowed for the user to check whether there have been periods of time where the person in the property has not moved, despite it being light in the room, indicating that a person was present but not able to move.

There was no clear difference between the two properties, with both having a similar number of detected anomalies. This is likely because of the similarity in the behaviours between the two properties. It's also unlikely that either of the people within the properties allowed for the properties to hit particularly uncomfortable levels, preventing the more severe anomalies from presenting themselves.

10 User Interface and Visualisation

This section aims to address the design decisions made when creating the app. This section specifically aims to address the research question of whether a system can be designed to provide information from the anomaly detection system in an easy to read way. We will discuss the design decisions made when designing the application. This section will also include an overview of the different pages that are included within the application.

10.1 Design

While there were no strict design guidelines provided by our partner company, there were applications that they had available to adhere the design to. Ensuring that the designs of the system that I created were in fitting with these applications was important. As well as sticking to the general Safehouse style, I also attempted to follow good practice when designing and implementing the rest of the systems visuals.

10.1.1 Front End

In the application Dash has been used for the front end. This is an open source Python framework created by the developers of the Plotly graphing library. According to the developers it is 'the most downloaded, trusted framework for building ML & data science web apps.' Dash is a Python based framework that allows for integration of both Python and HTML into the same web application, and is built on top of Plotly.js, React and Flask.

As Dash is built on top of React it is simple to make an application that is highly reactive. What this means is that the application will adapt the proportions of the elements on screen in order to accommodate the size of the screen, or window, that it is contained within. This allows for the application to be run on a multitude of different devices, such as desktop computers or mobile phones, without the need for the developer to explicitly accommodate these devices.

Both web and mobile app were unfeasible for the time that was available, but designing the application using a reactive framework ensures that a mobile application can be created later on if needed.

Bootstrap was also integrated into the application using the Dash Bootstrap Components package. Bootstrap is a responsive mobile first web framework. It is the most popular of its kind and has a range of components and tools to ensure that any website using it correctly is highly responsive. Also included are basic icons and a CSS sheet that can be built upon and customised to the users needs.

The primary reason that Bootstrap was implemented was due to its ability to adapt to mobile devices, as that's a potential use case I looked to explore. Another benefit of Bootstrap is that it's a popular framework and used frequently by popular websites, so it's likely that the user of the system will be familiar with the components that I am using from the library, such as the navigation bar and information cards. This will help to reduce the amount of training time required to get the user to a competent level with the system.

The charts used within the system are rendered using the Plotly Python library. Plotly is a free and open source graphing library that makes it possible to rapidly develop interactive charts and diagrams. The main advantage of Plotly over other graphing software is the high level of customisation that is available, as well as its ability to be easily integrated into a user friendly front end using the Dash framework.

As Plotly is a Python library it's able to directly interact with and have data passed through to it. This is important for allowing the user to interact with the system. Particularly when the user wants to test the system on different data sets or to make simple adjustments, such as to the format of the displayed data.

10.1.2 Charts

The charts are created and displayed using the Plotly graphing library. Line charts are used in order to map the key values of the sensors. These can accommodate the change in the sensors value over time and map it as such on the chart. This will make it possible to see how the sensor value has changed over the chosen duration of time. Line charts also make it easy to see the highs and lows in the data as these will show as the highs and lows of the chart.

Detected anomalies are mapped onto the chart as scatter points. Each different type of anomaly is automatically assigned its own colour to differentiate them from each other. This colour mapping is handled by Plotly and will adjust based on the number of different anomaly types that are input.

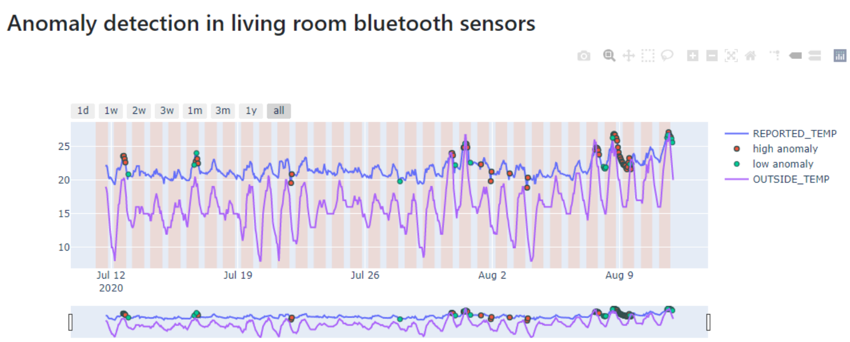


Figure 39: Basic chart

10.2 Pages

The user selects the property that they would like to view by selecting from the drop down in the top left of the page. This drop down will show any properties that have been allocated to the user. The currently selected property is shown in the top left of the page and will be used by the system when attempting to view readings from the different rooms.

Figure 40: Navigation panel

Each of the rooms in the two properties have been allocated their own page in the application. These pages remain consistent when either property has been selected. Different pages are available for the living room, bedroom and kitchen. In the event that sensors were installed in other rooms in the property then extra pages to accommodate them would become available. As the testing properties have both standard and bluetooth sensor types in the living room, an extra page has been included to allow the user to monitor either type.

Separate to these pages is a dashboard. This page contains localised information for every sensor in all of the property's rooms. If the user wants to monitor every room and sensor from one centralised location, they can do that on this page. This dashboard will be elaborated on in a later part of this section.

After going to Safehouse for feedback we found that at times the user was confused about the purpose of each of the extra conditions. This was primarily due to the short titles for the extra conditions, as longer titles would not be able to fit on the chart without taking up excessive space. To rectify this issue short descriptions for each of the extra conditions have been provided in the pages where those details are relevant.

10.3 Customisation

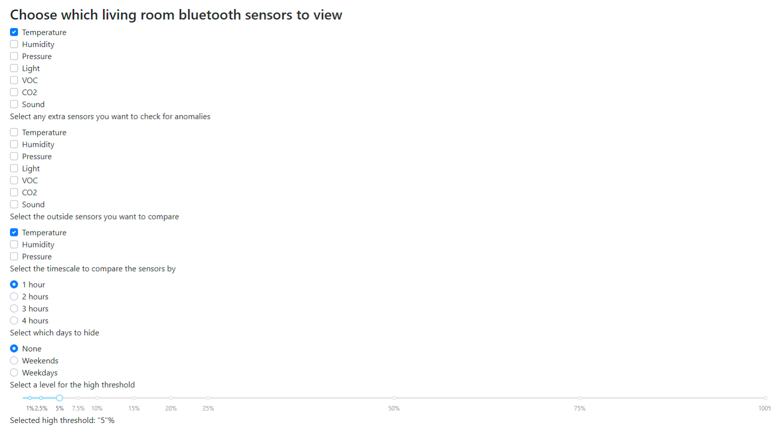


Figure 41: Basic page customisation

A large range of customisation options are available to the user. These customisation options present allow for a great deal of control over the system, both in the way that anomalies are detected as well as how they're displayed.

Each of the pages has different customisation options in order to accommodate their different use cases. The most basic of these differences is the sensor types that are able to be selected for the chart. Where appropriate, multiple sensors are able to be displayed on a single chart to facilitate easier comparison between their readings.

Where appropriate outside sensors can also be displayed. This is useful for comparing where dips in sensors, such as temperature, may be due to things changing outside of the property. In some cases there may be separate anomalies that occur because of what's happening outside, such as when the humidity indoors and outside cross.

By default each of the data sets is scaled to the nearest hour, for readings such as temperature or humidity this is the mean for that time, while in counting sensors like motion this is the total number of occurrences in that time period. This value can be easily changed in the interface so that another time scale can be used. This change not only alters what is displayed to the user but also the data fed to the algorithm, altering the results of the system itself.

The biggest change that can be made through the interface is to the two thresholds that are used by the system. These thresholds dictate which points are considered anomalies and which ones are not. The default thresholds are the top 5% for high and the top 10% for low. After several tests I found that these two thresholds provided the most relevant results for the majority of the sensors. However if the user wants to change this they're free to do so. This can allow the system to be changed so it's as sensitive as required by the user.

An extra option present on each of the pages is the ability to hide either weekdays or weekends from the charts. While this does not modify the actual results of the system, this makes it easier for the user to observe patterns in the results. This is particularly important as the behaviour of the occupant is likely to be significantly different on the weekends in comparison to weekdays.

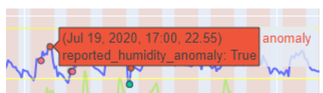


Figure 42: Detail page multiple charts

One of the unique features of the 'detail' page is that multiple sensors can be compared at the same time. While this is possible in the other charts by plotting all of the sensors on the same chart, this method can be more effective as it shows where other anomalies coincided at the same point on the chart. This is particularly useful when the user wants to compare anomalies across a large number of different sensor types, which may have values on a vastly different scale.

The 'multiple' page includes functionality to combine the results of several sensors into one. Testing has shown that using multiple relevant sensors can provide more accurate results than when a single sensor is used by itself. In this instance the user is able to select a main sensor along with one, or several, extra sensors. The main sensor is what will be predicted, while the extra sensors supplement that prediction.

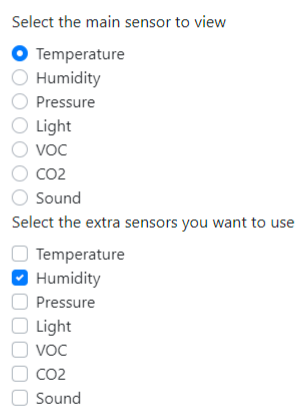


Figure 43: Multiple page customisation

10.3.1 Chart Customisation

Several options for customising the chart itself are also available. The main purpose of these options is to increase the readability of the chart and to prevent unnecessary information from cluttering the chart.

Results can be removed from the chart so that important details can be focused on. Both identified points and the lines used to represent values can be turned off. The chart size will resize to accommodate whatever data is being displayed. As can be seen, turning off the outside readings can greatly increase the visibility of what's going on indoors, at the expense of temporarily losing extra information.

Alongside the bottom of the charts is a scaled version of the chart. By using this tool the user is able to accurately select the time frame of data they want shown on the main chart. Not only can the user select a different period of time, but they're also able to adjust the length of this time to suit their needs.

Figure 44: Chart scale

The user is also able to interact directly with the charts. By simply highlighting a section using the mouse they can choose what data they want shown. Initially this highlighting allowed the user to zoom in on a specific point, however as this could be confusing when there are multiple sensors on the chart, this functionality was limited to allowing the user to highlight a specific amount of time. The chart will automatically zoom to an appropriate level for the chosen time, ensuring that the relevant sensor data is easy to see.

Another method to adjust the time frame of the chart is by using the buttons located at the top of each chart. These buttons scale the time frame based on which button has been pressed. Several useful and commonly used time frames are available.

For the sake of the test properties the buttons available are for a day, one or multiple weeks, one or multiple months, one year, or all of the data. An advantage that buttons have over manual adjustments is that a high level of precision is much easier to obtain and a specific amount of time can be easily selected, whereas manually adjusting the scale is susceptible to user error.

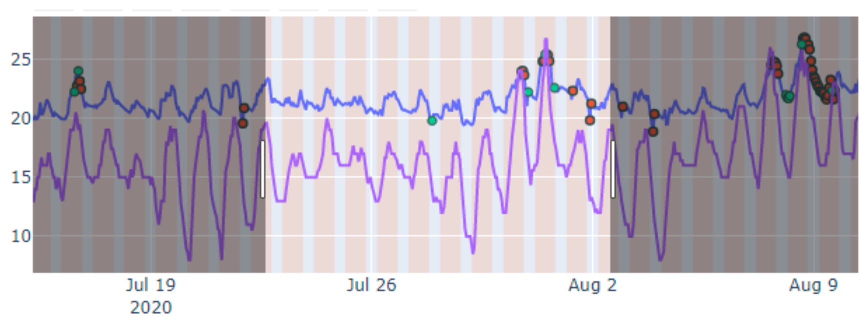


Figure 45: Chart highlighting

Figure 46: Date buttons

10.4 Dashboard UI

10.4.1 Prototype Design

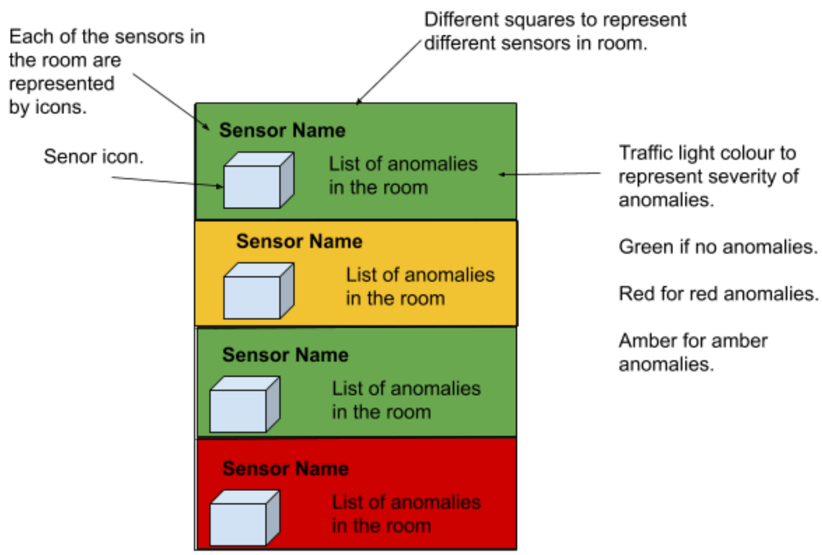


Figure 47: Dashboard Prototype

Before suggesting ideas to the Safehouse team I first devised a design prototype of how the dashboard would look. This would allow me to get early feedback before I proceeded to creating the actual dashboard.

The purpose of the dashboard is to convey all relevant information from the latest hour in a single location. The main advantage of this is the ability for the user to see at a glance whether any immediate action is needed, as well as whether it's worth investigating activity in a given property any further.

The system is separated into card blocks for each of the rooms, that are themselves separated into different sections for the sensors contained within. To represent the sensor types relevant iconography and text are used. While text is important for providing detailed information, icons allow the user to quickly find the sensor that they are looking for. Initially I intended for the entire sections to change colour based on which anomalies were detected in the room, however early testing showed that this made it too difficult to view the text in some instances. Instead only the colour of the sensor icon now changes. This colour changes based on which of the traffic light anomalies have been detected in the sensor.

Other anomalies detected in the system are listed alongside the sensor icon. As the majority of these anomaly types were deemed less immediately problematic than the traffic light ones it made sense to place them in a simple text view.

10.4.2 Final Design

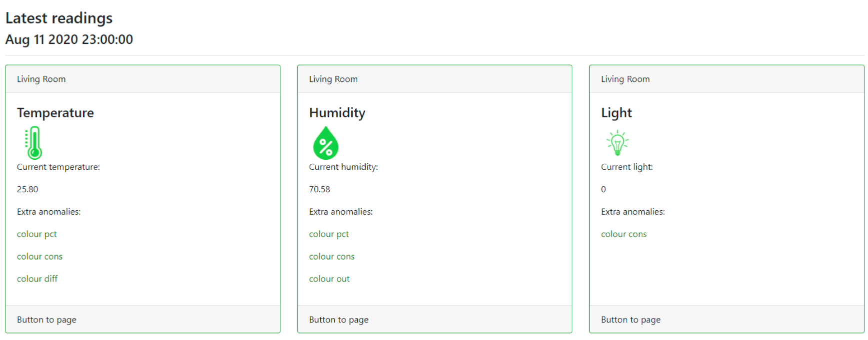


Figure 48: Dashboard Final

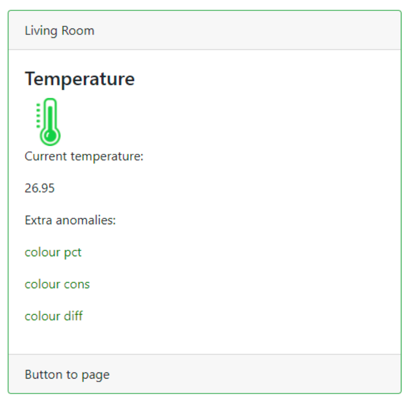


Figure 49: Dashboard card green

As in the prototype, the dashboard interface is split into sections based on the number of rooms being monitored in the property. These rooms are then split into several cards, shown in the fig below, that contain the relevant information on their given sensor, such as the latest reading and which anomalies were detected.

Indicator icons at the top of the card change colour based on whether a red, amber or green anomaly is detected in the sensor. This change is also reflected in the border of the card which also changes colour to match the icon. In cases where the traffic light system is not possible for the sensor type then green or red is used based on whether a standard anomaly was found.

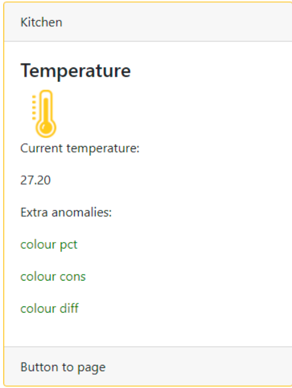


Figure 50: Dashboard card amber

Below these are details on other anomalies that were detected for this hour. These can help to give the user a better idea of what's going on in the property.

At the bottom of the card is a link to the corresponding page for that sensor that displays the data in a chart over a longer period of time. This is useful for monitoring the events that lead up to an anomaly occurrence and can be used by the user to help decide whether they want to act on the anomaly.

10.4.3 Usage

The user of the dashboard is expected to be a caregiver for a patient. This caregiver is envisioned to be an NHS worker or an independent carer. At present the dashboard is designed for one carer to monitor a limited number of patients. The dashboard provides a simple overview of the rooms and sensors for an individual patient, however the carer can navigate to the dashboards for other patients in their care by using the drop down menu in the navigation bar.

While more patients could be added to the system, this could quickly become overwhelming as the carer would need to manually navigate to the pages of each of the patients. For this reason the system is not ideally suited to the management of more than a select few patients. Future modifications, such as an alert system that indicates when an anomaly is detected in a patient not currently being viewed, could be implemented to allow for easier patient management.

Patients would also be able to view their own property through the system. In this instance the patient would only have access to their own sensors, and would not be able to navigate to another patients details through the dashboard.

10.4.4 Privacy

Privacy is a large concern for this type of application. While personal details about the patients are omitted from the application, it still relies on data obtained from the property that they're living in. This is sensitive data and will need to be treated in a secure manner. Were this application to be adopted for real use extra security would need to be provided to ensure that only authorised users can access the data for their patients. This would be implemented in the form of a login system that would prevent access to the dashboard or any other pages in the application unless the correct access rights are provided.

The application will also need to comply with local regulations, such as GDPR. This will include requirements, such as the need to host the data securely within the country and allowing the patient to have their information deleted on request. Patients would also need to be made aware of the data that is being tracked, how long it will be held for and where it will be held.

10.5 Conclusion

The feedback from the Safehouse team indicates that the design of the application was able to provide an effective interface for interacting with the data provided by the machine learning system. Over several iterations the dashboard was improved using their feedback to provide a detailed overview of the sensors from a single page. Each of the individual sensor pages allow for a range of customisation options to optimise the results of the system for the users needs.

There are some improvements that would be necessary for the application to go further than this project. Primarily around ensuring that the application is secure and easily scalable to more users. As detailed above this could be achieved through a login system that would limit which users are able to access the system.

11 Discussion

Overall the system was able to detect anomalies in different sensors with varying degrees of success. The customisation present allowed for the user to increase, or decrease, how strict the anomaly detection was. Allowing the number of anomalies detected to be fine tuned. While the extra conditions helped to highlight the most important anomalies that were detected, as well as providing a potential reason behind these anomalies.

11.1 Anomaly Detection

Humans act in an erratic and irregular way which can be difficult for machine learning models to learn due to the lack of clear patterns. This can be seen in the results of using the system on specific sensors such as temperature and humidity, that are only moderately impacted by the persons actions and behaviours, and may follow a set daily pattern. In comparison the system doesn't work as effectively on the motion and light. As these two sensors are directly impacted by the occupants random habits, such as what room they walked into and whether a light was turned on or not.

World events also had a large impact on the results of the system. Midway through the data gathering stage Covid-19 was reported in the United Kingdom, leading to lock downs and general changes to the populations behaviour. One of the changes that occurred was that people were told to remain in doors more often. This would have an impact on the sensor readings, particularly at times where usually the occupant would be out of the house, but were not because of Covid-19. Changes in behaviour like this make it more difficult for the models to make accurate predictions, as what may have been correct before Covid-19 may not be during the Covid-19 period. Requirements for the study also changed over time as initially it was unknown which sensors would be available for the project. Therefore the initial system was unable to be designed with specific sensors in mind. Instead over the course of the project the system needed to be adjusted to match the data that was provided.

Of the single sensor models used the AutoEncoder with an LSTM layer performed the best. As a model it was able to effectively detectl anomalies in the data set when compared to the two alternative models tested, as it was quicker while remaining accurate in its detection.

The system worked best when a combination of linked sensors were utilised together. Most notably in the case of temperature combined with humidity where the system worked very effectively and was able to make predictions with a high degree of accuracy. The major drawback of using the combination of sensors is the need for a significant amount of data, although this is slightly negated as data from multiple sensor sources on the same hub can be used together. Another issue with this method is the amount of time that it takes for the system to process data from multiple sensors together. This is particularly problematic when the predictions need to be available instantly to the person monitoring the occupant.

The difference in the results of the two different properties was relatively minimal. An interesting difference between the two was in cases where the person in one property followed a more clearly set pattern than the other. This is particularly visible in the light sensors where we can see that the lights came on at virtually the same time each day in the second property, but was on at seemingly random intervals in the other. Surprisingly this did not have a significant impact on the results of the machine learning models as both had similar RMSE, MAE and R2 values for this sensor.

Another notable difference between the two properties was the large variance in the humidity between the two. This was the only environmental sensor where a large difference was found between the results from each of the properties. This may indicate that there is a higher degree of randomness when it comes to humidity in comparison to other factors, or that the one property was better insulated than the other.

11.2 Extra Conditions

Extra conditions were useful for optimising the system once it was deemed by the Safehouse team to be able to detect anomalies with a good level of success. In particular the traffic light system worked well for reducing the number of false positive anomalies detected by the system. This limited the anomalies detected to those that are within set bounds unique for each sensor type.

Different boundary levels show the expected severity of any detected anomalies. Anomalies detected within the red boundary are severe and likely require immediate action, while amber anomalies are not as severe and may not require immediate action to remedy. On the other hand any anomalies within the green boundary are ignored and no longer marked on the chart. These factors can be useful for the user when deciding whether they need to act on the detected anomalies.

One issue with the traffic light system is that sharp drops that remain within these bounds will not be detected, and that the bounds may not always be accurate for every property or person. Some occupants may prefer the temperature, humidity or light to remain within a different range for them to feel comfortable at home. These are factors that could not be accounted for as each person would need to be polled for their ideal comfort level.

Other conditions that were added, such as the comparison with outside values, helped to define the cause of the anomalies that were detected, or to detect strange activities that would not usually be classified by the system. This helped to ensure that the system provided a comprehensive view of all potential anomalies that may be occurring in the property. Many of these requirements were not dependent on the machine learning part of the system, and highlight the fact that many anomaly types can be detected without needing to develop a complex system. In some cases these simple checks performed as effectively, or more effectively, than the main system as they were able to detect anomalies that were missed or filter out potential false anomalies.

Conditions unique to the motion and light were effective in determining which anomalies it was worth the user investigating. Initially the system attempted to detect movement when the room is dark, however because people don't generally turn on their light in the daytime this resulted in a number of anomalies that we deducted were false positives. To refine this only anomalies of this kind detected at night, as defined by the local day night cycle, were marked as anomalies on the chart. This cut down on the number of points in time that the user needed to be aware of, while also limiting the anomalies to those that would be the most dangerous, as the occupant is more likely to be in danger when moving in the dark night, rather than in the day.

11.3 Measuring Effectiveness

None of the measurement criteria used by themselves were able to accurately assess the effectiveness of the model, but by combining the results of different tests a general idea of the effectiveness of the systems could be determined. Unfortunately due to the random nature of human activity the R2 score was not as useful as first hoped, instead more subjective measures such as MAE and RMSE were relied upon. Combining these methods allowed for a fair comparison between the different model types, to find which performed the best on the data that was available.

For a more detailed analysis of the models effectiveness the feedback from our partner company was used. Not only did this provide feedback on the effectiveness of the model, but it also gave some insight into the reason that the anomalies were detected in the first place.

11.4 User Interface

The interface effectively displays the results on the different sensors. One, or multiple sensors, in any of the properties can be plotted to chart and viewed by the user. Individual sensor pages gave a detailed overview of the events that led up to the detected anomalies. Allowing the user to track the data from when tracking began to view the trends for themselves, before choosing to agree or disagree with the assessment of the anomaly detection system.

Extensive customisation was available to the user on every page, allowing them to change each aspect of the system to their liking. This is particularly useful for limiting the amount of sensors shown on the charts, which is important for clearly analysing activities in the property. Charts will auto scale to account for the amount of data on display, but the user is also able to highlight a select amount of data to exclusively display. Over the course of the project this functionality was improved to provide a more convenient way to choose what data was on display. A selection of buttons were made available that could be pressed to select a specific period of time. Buttons were available for the common time frames that may be needed, and the data on display would be limited to such. As well as this the user can manually highlight portions of the chart to zoom in on them. This is useful for analysing the change in data that caused for an anomaly to occur.

Options were also present to change how the system dealt with the data. The user can choose which time scale, to the hour, that the data will be formatted to. As well as the time frame the user is able to select the thresholds used for classification, allowing them to make the system as sensitive or strict as they want. Both of these options will vastly change the results of the system and which points are classified as anomalies. Effectively accommodating the needs of the user. We found that the ability to change the threshold was also particularly useful during the testing process, as we could minor changes to the threshold and then observe the results.

The dashboard was able to accurately convey an overview of the property at any given time. A quick glance at the dashboard would allow for the user to quickly identify any sensors within the property that may require immediate action. In the event that a recent anomaly was detected the user could then click into that sensors individual page for more information. This was an effective way of providing the immediately relevant info quickly, while allowing for the user to obtain more information if required.

To ensure that the details presented on the dashboard were clear a combination of icons, colours and words were used. Icons to represent each of the sensor types were present in each of the room cards and would change colour based on whether a recent anomaly was detected. Red, amber or green were chosen for this purpose to give an obvious indication of the state of the room using a universally recognised colour scheme for danger. In instances where more details on the anomalies found would be needed a list of identified anomalies were also provided to the side of the sensor icons. These were not colour coded as they were not deemed as immediately important as the traffic light anomalies and multiple colours could end up reducing the readability of the rest of the page.

An intrinsic limitation of the study is that the environment may not always accurately reflect the actions of the occupant. Outside factors will have a large impact on the readings of many of the sensors. One way we attempted to alleviate this problem was by displaying the readings of similar sensors outdoor on the same chart, thus allowing the user to compare the readings inside and outside, and identify any potential corresponding trends. This was possible in sensors where we had outside reading equivalents, such as temperature and humidity.

Overall the Safehouse team deemed in the feedback sessions that the dashboard interface was effective in quickly and efficiently communicating the most important information from the system. The only notable disadvantages of the dashboard are that it takes a longer period of time to load than the individual pages, as it needs to process each sensor in every room, and that accessing more detailed information requires the user to navigate to a separate page altogether.

11.5 User Comfort

While the idea of ambient assisted living is becoming more popular, there are still several factors holding it back, as discussed by (Memon et al. 2014) these factors ranged in level of severity and importance. Multiple studies as highlighted in section 5.1 highlighted the ways, and reasons, that people may feel uncomfortable with being monitored, primarily related to their independence and privacy. The study by (Rashidi and Mihailidis 2013) found that it was very important that the sensors were as comfortable as possible for the participants. They also raised concerns about the security of the participants' data, and the importance of ensuring that the proper procedures are taken to ensure that it is as secure as possible. While (Maan and Gunawardana 2018) highlighted participants concerns around their own technology skills hindering the system, or that it would result in less human contact. These are concerns that we hope to have addressed by making our system as unobtrusive as possible, resulting in fewer users feeling that their privacy is being invaded, or that they are being monitored at all times. The person being cared for will also not need to be technologically proficient, as they do not need to directly interact with the system in anyway.

11.6 Property Data

This study had a number of limitations that effected what could be done. The most notable limitation was the amount of data accessible. Only two properties were being tracked during this study, and the amount of time that the properties were monitored was also limited by external factors. This led to a reduced amount of data compared to what was initially expected for the project.

The small sample size of different properties also limited the number of times that the models could be tested, making it difficult to assess whether unusual results in the readings were because of issues in the model, or external factors that may be present in these specific properties alone. This also introduced the issue of being unable to accurately detect seasonality within the data. The range of values detected by many of the environmental sensors, such as temperature, would be vastly different during different seasons. While the model created has tried to account for differences between month, the lack of data makes it impossible to accurately model how different whole seasons would be.

The data was of mixed quality, and while the data did accurately represent the properties of two people within our target demographic, neither of the environments were completely controlled. The presence of other people, or pets, entering the household could not be monitored, potentially resulting in unusual sensor readings as unexpected people entered the properties.

While this was intentional for ensuring that the system can work in realistic scenarios, it made it difficult to assess the accuracy of the system.

While we aimed to have sensors in similar locations in the two properties, the difference in those properties meant that they could not be in identical locations in the two. Proximity to doors, or windows, when opened would have a large impact on several environmental readings, such as temperature, humidity and light. For these reasons the properties are not able to be directly compared on the actual values of the sensors, instead trends, spikes in data and instances of anomalies detected in the properties are compared.

Neither of the properties used were closely monitored meaning we had no exact record of the activities taking place within the property. Because of this we were unable to classify anomalies as specific activities and were limited in the type of models that could be used. This caused further problems as it left us without the ability to mark identified anomalies as true or false, instead we needed to measure the anomalies using more subjective measurement methods. Eventually a threshold system was settled on. This threshold system would use the training data to dictate what are suitable ranges for the thresholds to be within. Ultimately this appeared effective, with our industry partners agreeing with the assessment of the system for the most part. The extra conditions added also helped to ensure that the results that came from the system were valid anomalies, and not false positives.

12 Conclusion

As detailed in the introduction of this paper we aimed to answer several research questions in this thesis. In this section we will discuss whether or not these research questions have been met, as well as any future work that may be needed for this project to take it further.

What is the most effective machine learning model for detecting anomalies in sensor data?

Each of the models tested had their own advantages and disadvantages. Overall the Autoencoder was the best solution for general anomaly detection as it provides consistent results, while being able to process data at a rapid pace. This is particularly important as in a real world setting the system would be hosted on a computer, or server, of a lower specification than of that used in this study. The amount of data used for the system is also likely to grow exponentially increasing the amount of processing required.

Of the models LSTM is the most effective at accurately predicting future values for the sensors, however the actual anomaly detection was not necessarily more effective as we account for the potential error when incorporating the thresholds. This model also took a considerable amount of time to train on new data, making it a potentially unviable solution for real world application, especially as the data set continues to grow.

Does a combination of different sensor types provide more accurate anomaly detection results than a single sensor solution?

As shown in section 8.4, using a combination of relevant sensors improved the quality of the prediction made significantly, with the combination of temperature and humidity working particularly well together. However, as with the univariate LSTM model, this also had the issue of taking a great deal of time to process the data. In the event that processing time was not an issue a multiple sensor approach may be the most effective for accurately predicting anomalies. This also had the downside that if the two sensors were not related in a clear way that the results of the combination of sensors would be less accurate than utilising a single sensor, such as in the case where temperature was used with light.

Can machine learning be used on environmental sensors in order to detect anomalies in human behaviour?

The threshold system allowed for the number of anomalies detected to be fine tuned by the user, giving control over how strict the system was at any given time. Combining these results with a carefully selected group of extra conditions has helped to refine the detected anomalies down to only the most important and likely anomalies, as well as detecting anomalies that may have been missed by the system otherwise. While this may need some tuning in the future, the team at Safehouse has indicated that these have identified the most important potential anomalies. The use of extra conditions, such as the traffic light system, has also allowed the system to narrow down the type of anomalies. Anomalies are now graded by severity and type, allowing for a carer to asses whether they agree with the systems output.

Can a system be designed to provide this information in an easy to read way?

The feedback provided by the Safehouse team indicates system is effective in presenting the detected anomalies to the user. This is primarily due to the way that the system was designed around their expert knowledge in the field, as detailed in section 10. Charts in the application are automatically adjusted to display anomalies in an effective manner, with colour coding and annotations helping to highlight important information. Even in situations where the user is not necessarily interested in anomaly detection, the system remains a useful tool for visualising the data over time. With the high level of customisation in the pages allowing the user to customise charts to their needs. Regular monitoring will be required to act on the anomalies once they have been detected. Carers will need to manually evaluate detected anomalies by reading through the sensor data using the charts provided to ensure that they agree with the systems evaluation.

Can environmental sensors be used in combination with automatic remote monitoring to allow vulnerable people to live alone?

I believe that this system shows that anomaly detection is a viable solution for helping to allow older members of the community to live at home, as it can provide a valuable tool to assist carers by acting as an early warning system and identifying potential problems within a property, without requiring them to be there themselves. Not only does this allow for single carers to monitor multiple people, but anomalies detected may also identify risk factors that have been missed by the carer. This can also allow for elderly, or otherwise vulnerable people, to live with some degree as independence as they will not need to be physically monitored at all times. Not only will this reduce the pressure on the carers, but may also increase the amount of privacy of the person being cared for, as the sensors we are monitoring for this project can be considered less intrusive than having someone physically present.

While the system is effective at detecting anomalies this alone is not enough to allow a person to live independently at home. There are still the risks of potential problems that are unable to be captured by the environmental data available. For these reasons future work is needed to test the system in a realistic scenario to identify these potential gaps.

Future Work

The system may be improved by more closely monitoring the properties to check whether the anomalies detected by the system match up with the problems that are present at the property. At present we rely upon the measures that we can records such as the RMSE and R2 and the subjective expert opinions of the Safehouse team. Making this change would require a participant to keep a detailed diary of events within the property which can then be marked and checked. Keeping track of events will also allow the system to be improved by marking anomalies as true positives or false positives in the data set, potentially allowing for extra classification in the future.

While we initially intended for the system to be designed for a carer monitoring people living at home, there is scope to expand the system in order to accommodate multiple users, properties or even care homes. This would need to be done in the user interface by allowing for the carer to view the sensors for multiple rooms in different properties on the same page. The current system could also be adapted so as to provide notifications to the carer when an anomaly is detected, rather than expecting the carer to be viewing the dashboard or individual page at any given time.

13 Bibliography

References

Rabiner, L. R. and B. H. Juang (1986). "An Introduction to Hidden Markov Models". In: *IEEE ASSP Magazine* 3.1, pp. 4–16. ISSN: 07407467. DOI: 10.1109/MASSP.1986.1165342.

Kohzadi, Nowrouz et al. (1996). "A comparison of artificial neural network and time series models for forecasting commodity prices". In: *Neurocomputing* 10.2, pp. 169–181. ISSN: 09252312. DOI: 10.1016/0925-2312(95)00020-8.

Yeh, Shu Chuan Jennifer and Sing Kai Lo (2004). "Living alone, social support, and feeling lonely among the elderly". In: *Social Behavior and Personality* 32.2, pp. 129–138. ISSN: 03012212. DOI: 10.2224/SBP.2004.32.2.129.

Willmott, Cort J. and Kenji Matsuura (Dec. 2005). "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance". In: *Climate Research* 30.1, pp. 79–82. ISSN: 0936577X. DOI: 10.3354/cr030079.

Hyndman, Rob J. and Anne B. Koehler (Oct. 2006). "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4, pp. 679–688. ISSN: 01692070. DOI: 10.1016/j.ijforecast.2006.03.001.

Ahmad, Mohd Yazed and Ananda Sanagavarapu Mohan (2009). "RFID reader localization using passive RFID tags". In: *APMC 2009 - Asia Pacific Microwave Conference 2009*. ISBN: 9781424428021. DOI: 10.1109/APMC.2009.5384152.

Hsu, Hui Huang and Chien Chen Chen (2010). "RFID-based human behavior modeling and anomaly detection for elderly care". In: *Mobile Information Systems*. ISSN: 1574017X. DOI: 10.3233/MIS-2010-0107.

Van Kasteren, T. L.M., G. Englebienne, and B. J.A. Kröse (2010). "An activity monitoring system for elderly care using generative and discriminative models". In: *Personal and Ubiquitous Computing*. ISSN: 16174909. DOI: 10.1007/s00779-009-0277-9.

Khashei, Mehdi and Mehdi Bijari (Mar. 2011). "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting". In: *Applied Soft Computing Journal*. Vol. 11. 2, pp. 2664–2675. DOI: 10.1016/j.asoc.2010.10.015.

Fukunaga, Ryuta et al. (Sept. 2012). “Living alone is associated with depression among the elderly in a rural community in Japan”. In: *Psychogeriatrics* 12.3, pp. 179–185. ISSN: 1479-8301. DOI: 10.1111/J.1479-8301.2012.00402.X. URL: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1479-8301.2012.00402.x%20https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1479-8301.2012.00402.x%20https://onlinelibrary.wiley.com/doi/10.1111/j.1479-8301.2012.00402.x.

Lotfi, Ahmad et al. (2012). “Smart homes for the elderly dementia sufferers: Identification and prediction of abnormal behaviour”. In: *Journal of Ambient Intelligence and Humanized Computing*. ISSN: 18685137. DOI: 10.1007/s12652-010-0043-x.

Novák, Marek, Miroslav Biňas, and František Jakab (2012). “Unobtrusive anomaly detection in presence of elderly in a smart-home environment”. In: *Proceedings of 9th International Conference, ELEKTRO 2012*, pp. 341–344. DOI: 10.1109/ELEKTRO.2012.6225617.

Wu, Jiaqing (2012). “Three-Dimensional Indoor RFID Localization System”. PhD thesis.

Rashidi, Parisa and Alex Mihailidis (2013). “A survey on ambient-assisted living tools for older adults”. In: *IEEE Journal of Biomedical and Health Informatics* 17.3, pp. 579–590. ISSN: 21682194. DOI: 10.1109/JBHI.2012.2234129.

Chai, T. and R. R. Draxler (June 2014). “Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature”. In: *Geoscientific Model Development* 7.3, pp. 1247–1250. ISSN: 19919603. DOI: 10.5194/gmd-7-1247-2014.

Memon, Mukhtiar et al. (2014). *Ambient Assisted Living healthcare frameworks, platforms, standards, and quality attributes*. DOI: 10.3390/s140304312.

Patel, Jigar et al. (Mar. 2015). “Predicting stock market index using fusion of machine learning techniques”. In: *Expert Systems with Applications* 42.4, pp. 2162–2172. ISSN: 0957-4174. DOI: 10.1016/J.ESWA.2014.10.031.

Tsukiyama, Toshifumi (2015). “In-home health monitoring system for solitary elderly”. In: *Procedia Computer Science*. DOI: 10.1016/j.procs.2015.08.338.

Wang, Lidong, Guanghui Wang, and Cheryl Ann Alexander (2015). “Machine Learning in Big Data”. In: *International Journal of Advances in Applied Sciences (IJAAS)* 4.4, pp. 117–123. ISSN: 2252-8814.

Zhu, Chun, Weihua Sheng, and Meiqin Liu (Oct. 2015). "Wearable Sensor-Based Behavioral Anomaly Detection in Smart Assisted Living Systems". In: *IEEE Transactions on Automation Science and Engineering* 12.4, pp. 1225–1234. ISSN: 15455955. DOI: 10.1109/TASE.2015.2474743.

Aran, Oya et al. (2016). "Anomaly detection in elderly daily behavior in ambient sensing environments". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. ISBN: 9783319468426. DOI: 10.1007/978-3-319-46843-3{_}4.

Dean, Alfred et al. (June 2016). "The Influence of Living Alone on Depression in Elderly Persons:" in: *http://dx.doi.org/10.1177/089826439200400101* 4.1, pp. 3–18. ISSN: 15526887. DOI: 10.1177/089826439200400101. URL: https://journals.sagepub.com/doi/abs/10.1177/089826439200400101.

Myttenaere, Arnaud de et al. (June 2016). "Mean Absolute Percentage Error for regression models". In: *Neurocomputing* 192, pp. 38–48. ISSN: 18728286. DOI: 10.1016/j.neucom.2015.12.114.

Nishida, Yoshihumi et al. (2016). "Handrail-shaped IoT sensor for long-term monitoring of the mobility in elderly people". In: *Advances in Intelligent Systems and Computing*. Vol. 500. Springer Verlag, pp. 631–638. ISBN: 9783319419619. DOI: 10.1007/978-3-319-41962-6{_}56.

Science, Government Office for (2016). *Future of an Ageing Population*. Tech. rep., p. 124. URL: https://www.ageing.ox.ac.uk/files/Future_of_Ageing_Report.pdf.

Arifoglu, Damla and Abdelhamid Bouchachia (2017). "Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks". In: *Procedia Computer Science*. DOI: 10.1016/j.procs.2017.06.121.

Ewbank, Leo, James Thompson, and Helen McKenna (2017). *NHS hospital bed numbers: past, present, future*. DOI: 10.1186/2110-5820-3-37.

Yassine, Abdulsalam, Shailendra Singh, and Atif Alamri (2017). "Mining Human Activity Patterns from Smart Home Big Data for Health Care Applications". In: *IEEE Access* 5. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2719921.

Zhou, Lina et al. (May 2017). "Machine learning on big data: Opportunities and challenges". In: *Neurocomputing* 237, pp. 350–361. ISSN: 0925-2312. DOI: 10.1016/J.NEUCOM.2017.01.026.

Frías-Paredes, Laura et al. (Apr. 2018). "Dynamic mean absolute error as new measure for assessing forecasting errors". In: *Energy Conversion* *and Management* 162, pp. 176–188. ISSN: 01968904. DOI: 10.1016/j.enconman.2018.02.030.

Maan, Chetna and Upul Gunawardana (Jan. 2018). "Barriers in acceptance of Ambient Assisted Living technologies among older Australians". In: *2017 IEEE Life Sciences Conference, LSC 2017*. Vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., pp. 222–225. ISBN: 9781538610305. DOI: 10.1109/LSC.2017.8268183.

Hosseini, Anahita and Majid Sarrafzadeh (May 2019). "Unsupervised prediction of negative health events ahead of time". In: *2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings*. Institute of Electrical and Electronics Engineers Inc. ISBN: 9781728108483. DOI: 10.1109/BHI.2019.8834550.

Kieu, Tung et al. (2019). "Outlier detection for time series with recurrent autoencoder ensembles". In: *IJCAI International Joint Conference on Artificial Intelligence*. Vol. 2019-August. DOI: 10.24963/ijcai.2019/378.

Sagheer, Alaa and Mostafa Kotb (2019). "Unsupervised Pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate Time Series Forecasting Problems". In: *Scientific Reports* 9.1. ISSN: 20452322. DOI: 10.1038/s41598-019-55320-6.

Siami-Namini, Sima, Neda Tavakoli, and Akbar Siami Namin (Jan. 2019). "A Comparison of ARIMA and LSTM in Forecasting Time Series". In: *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*. Institute of Electrical and Electronics Engineers Inc., pp. 1394–1401. ISBN: 9781538668047. DOI: 10.1109/ICMLA.2018.00227.

Yamak, Peter T., Li Yujian, and Pius K. Gadosey (2019). "A comparison between ARIMA, LSTM, and GRU for time series forecasting". In: *ACM International Conference Proceeding Series*. DOI: 10.1145/3377713.3377722.

Deep, Samundra et al. (Jan. 2020). "A Survey on Anomalous Behavior Detection for Elderly Care Using Dense-Sensing Networks". In: *IEEE Communications Surveys and Tutorials* 22.1, pp. 352–370. ISSN: 1553877X. DOI: 10.1109/COMST.2019.2948204.

Living longer - Office for National Statistics (n.d.). URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglongerhowourpopulationischangingandwhyitmatters/2019-03-15.

Overview of the UK population - Office for National Statistics (n.d.). URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/august2019.

Shen, Shunrong, Haomiao Jiang, and Tongda Zhang (n.d.). "Stock Market Forecasting Using Machine Learning Algorithms". In: ().

14 Appendix

14.1 Extra tests

Batch	RMSE	R2
64	0.285	0.928
128	0.271	0.934
256	0.263	0.938
500	0.275	0.933

Table 58: The results of changing the batch size of the model when used on the Temperature Bluetooth data from the first property with a 9 to 1 split

Epochs	RMSE	R2
64	0.254	0.942
128	0.263	0.938
256	0.266	0.937
500	0.273	0.934

Table 59: The results of changing the epochs of the model when used on the Temperature Bluetooth data from the first property with a 9 to 1 split

14.2 Data

Figure 51: Property 1 Bluetooth data

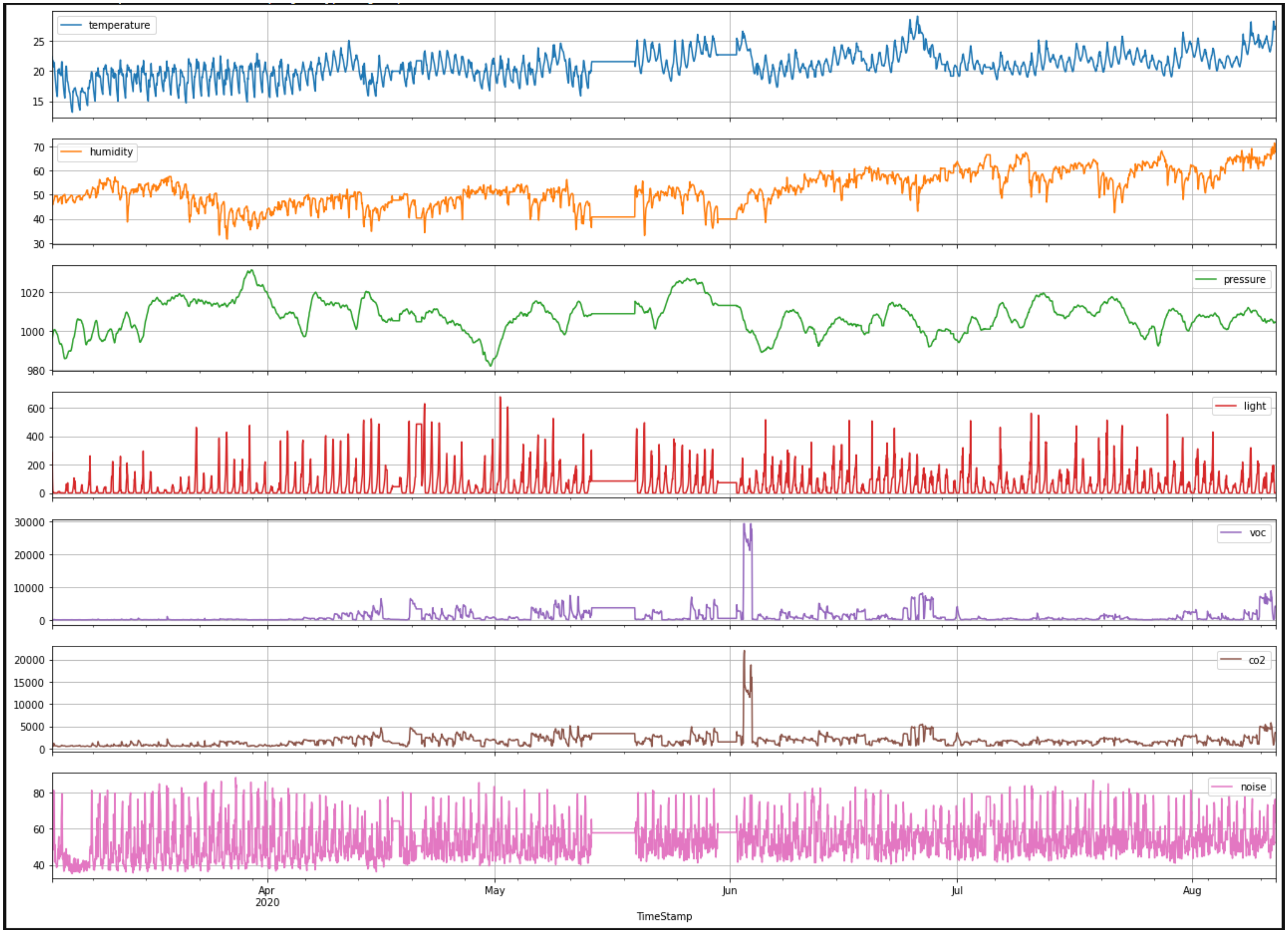
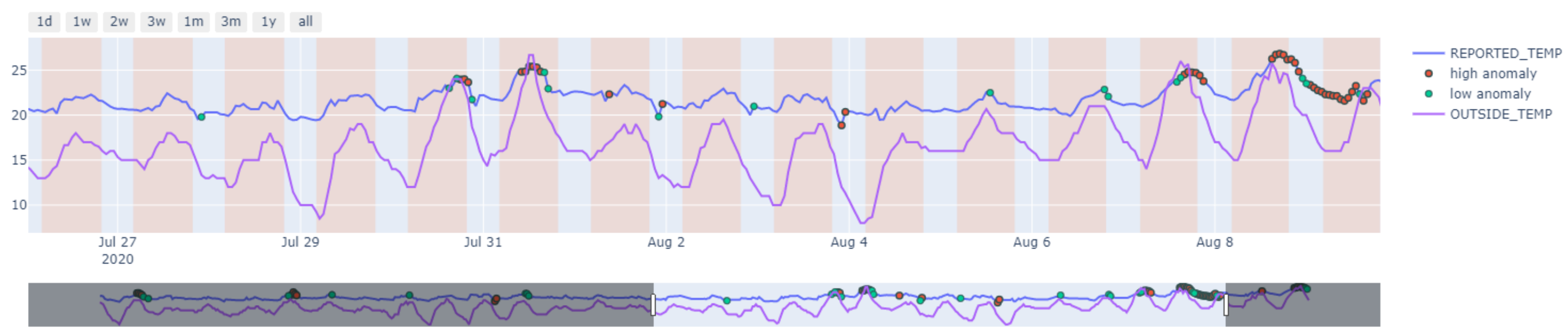


Figure 52: Property 2 Bluetooth data

14.3 Results

Celsius



Date

Figure 53: Regular anomaly detection in property 1 temperature

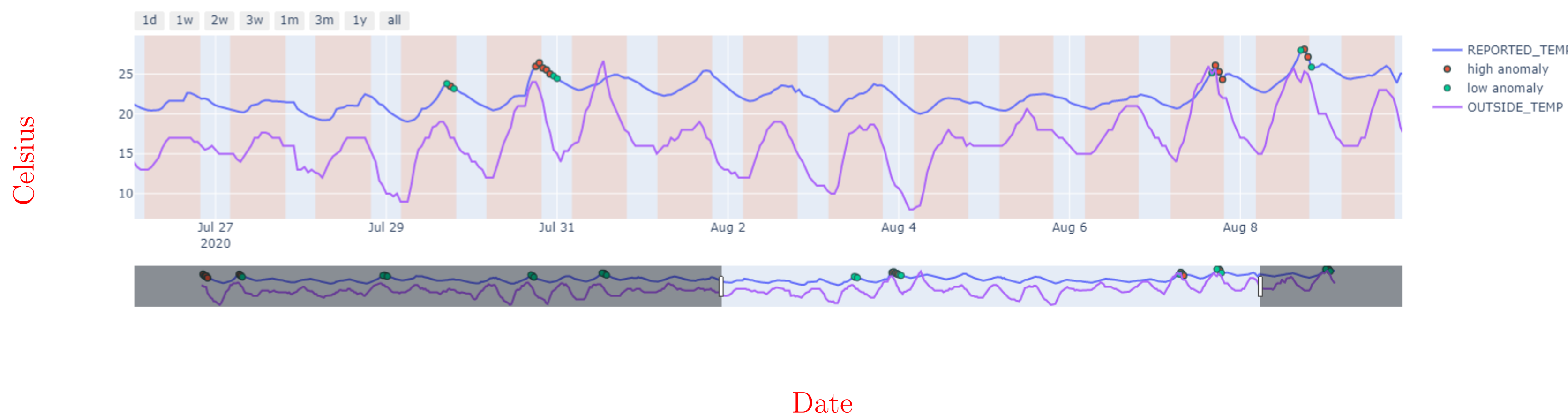
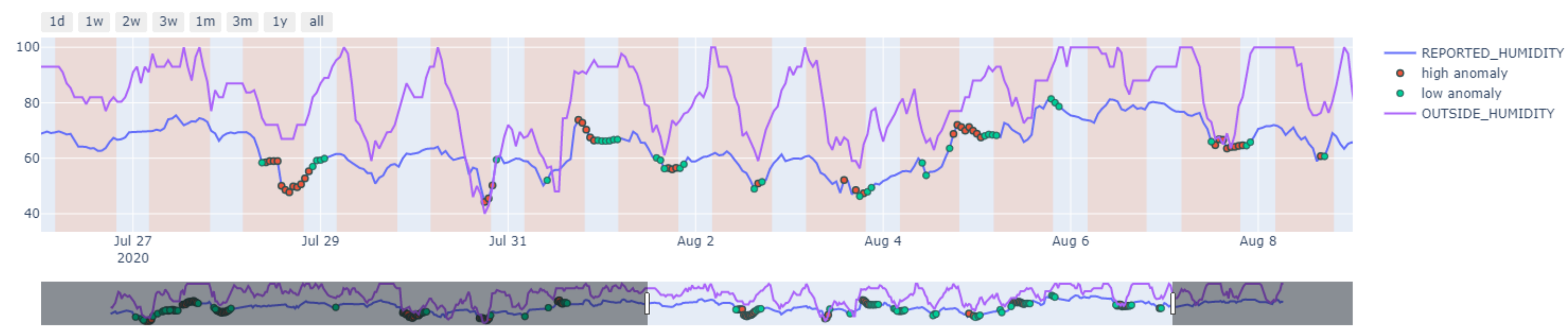


Figure 54: Regular anomaly detection in property 2 temperature

Humidity Percentage



Date

Figure 55: Regular anomaly detection in property 1 humidity

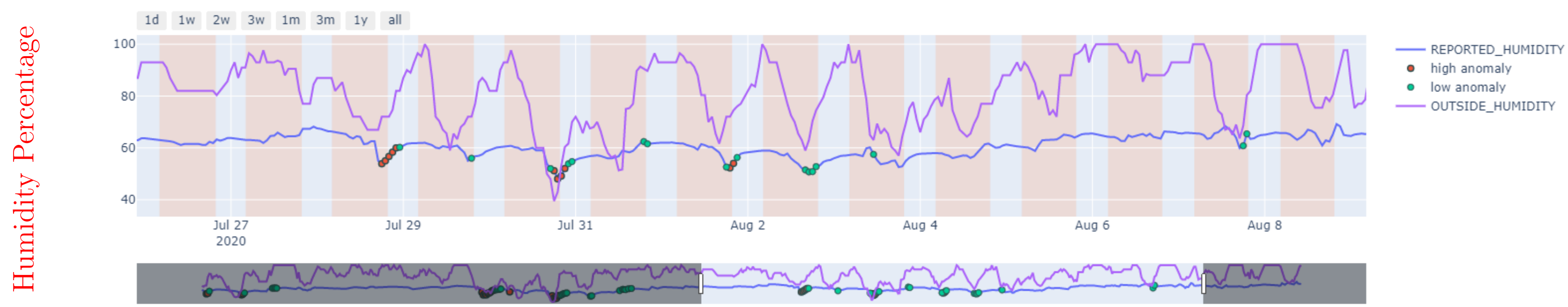


Figure 56: Regular anomaly detection in property 2 humidity

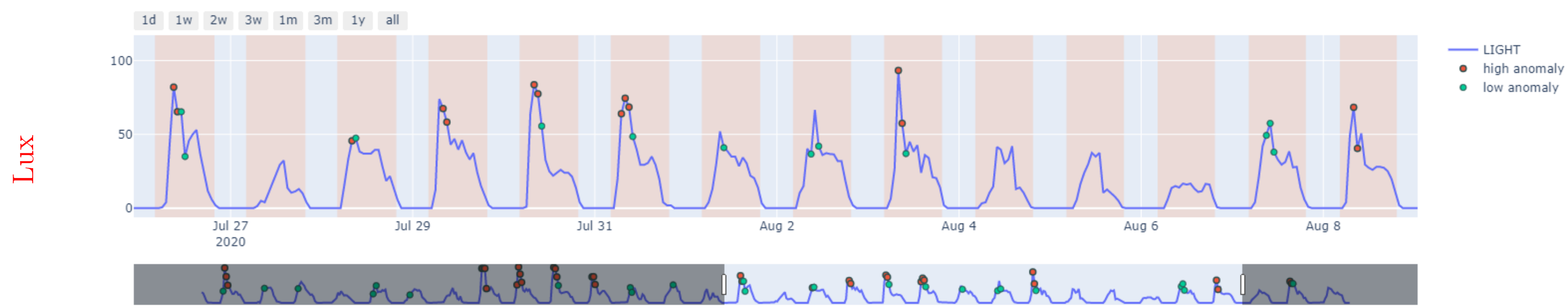


Figure 57: Regular anomaly detection in property 1 light

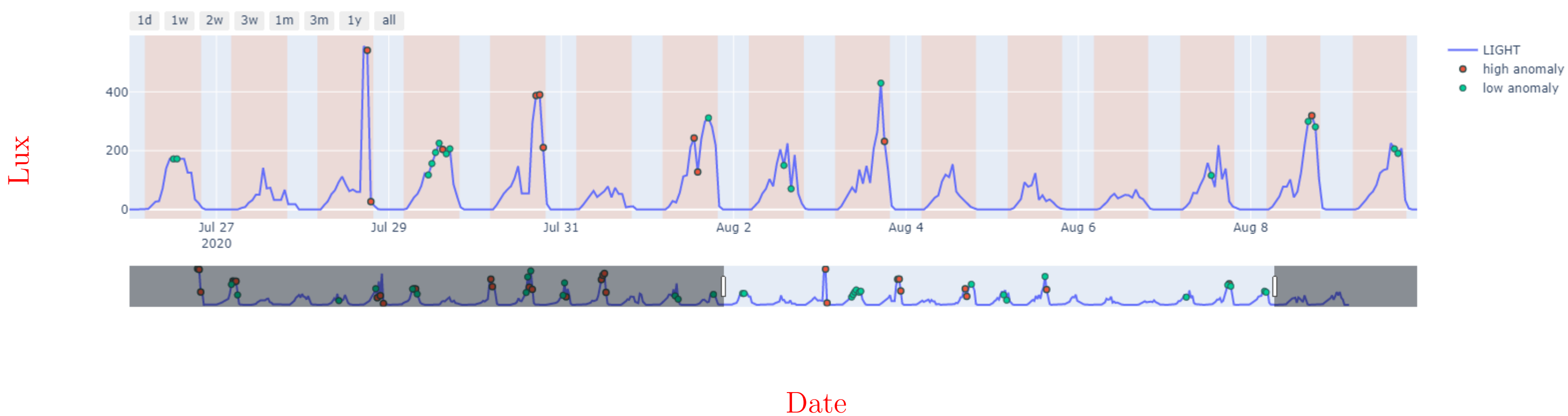
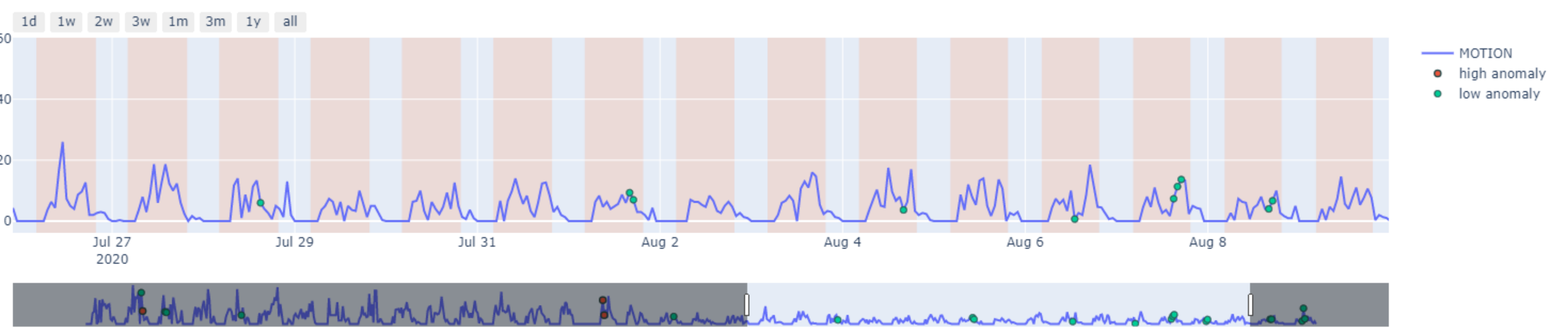


Figure 58: Regular anomaly detection in property 2 light

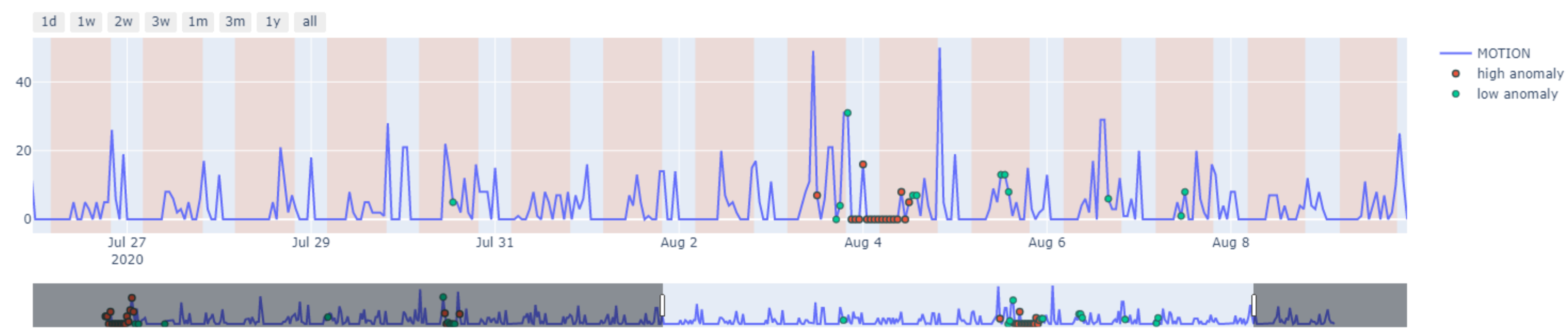
Motion Recorded



Date

Figure 59: Regular anomaly detection in property 1 motion

Motion Recorded



Date

Figure 60: Regular anomaly detection in property 2 motion

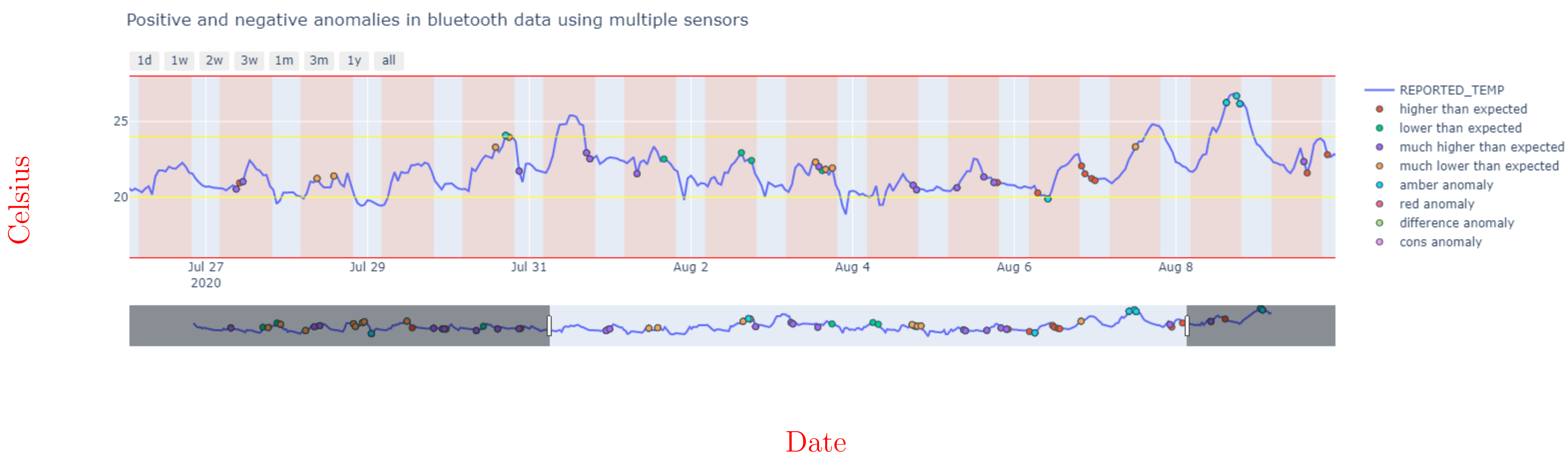


Figure 61: Anomaly detection in property 1 temperature supplemented by humidity

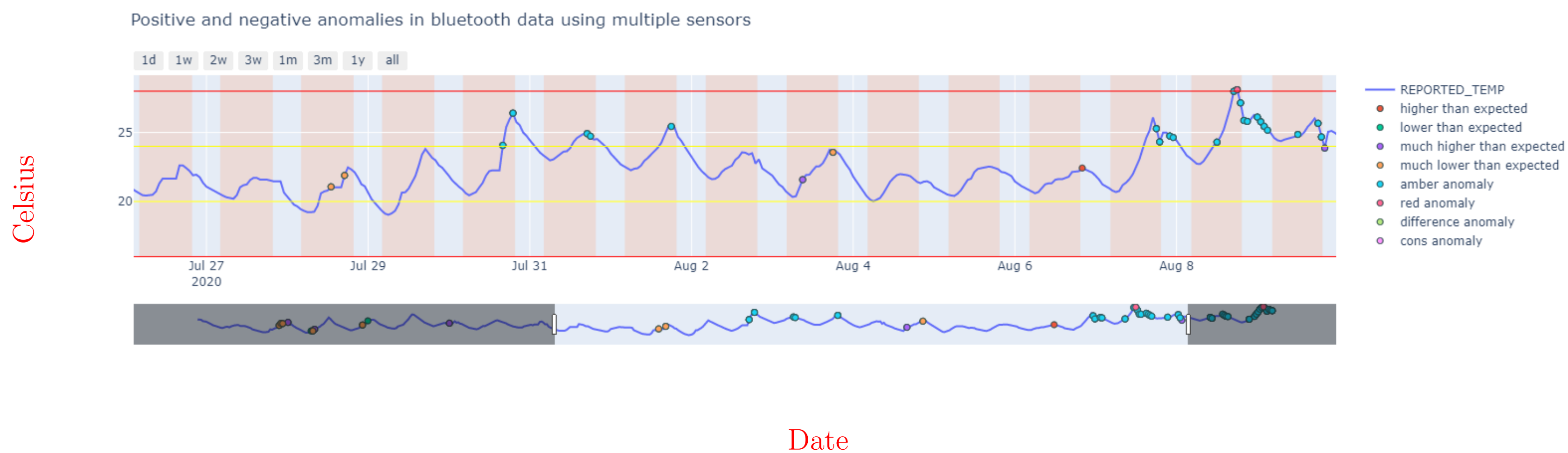


Figure 62: Anomaly detection in property 2 temperature supplemented by humidity

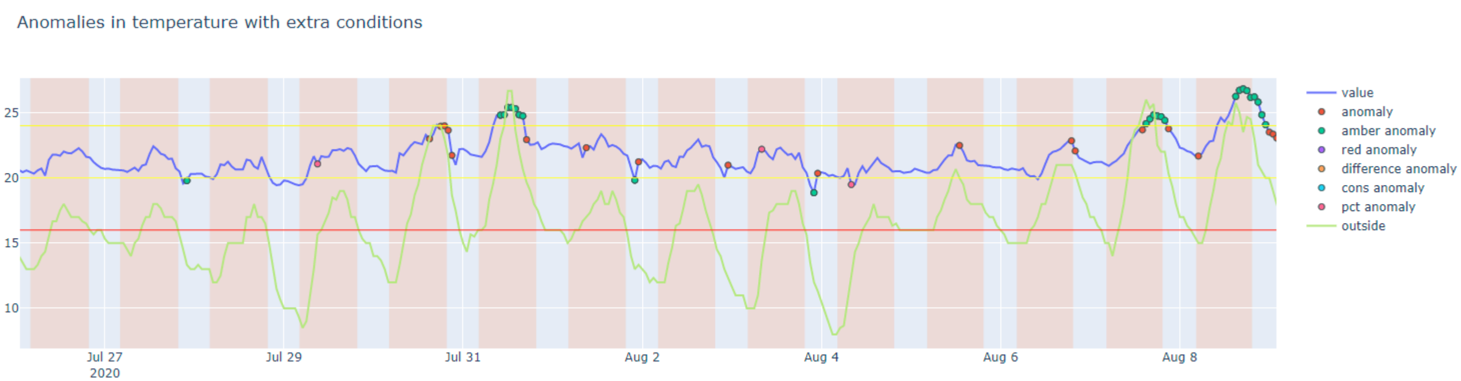


Figure 63: Property 1 Temperature with all conditions

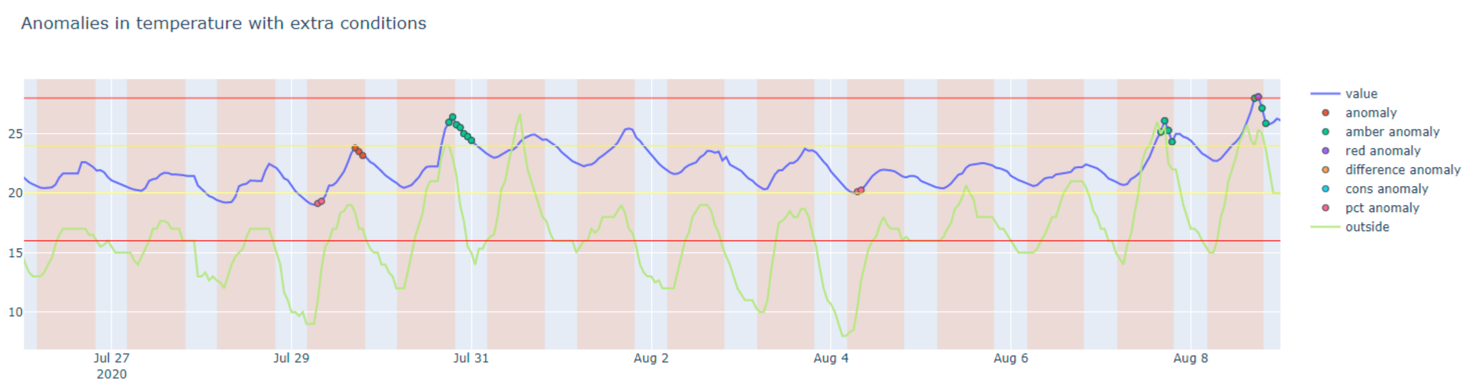


Figure 64: Property 2 Temperature with all conditions

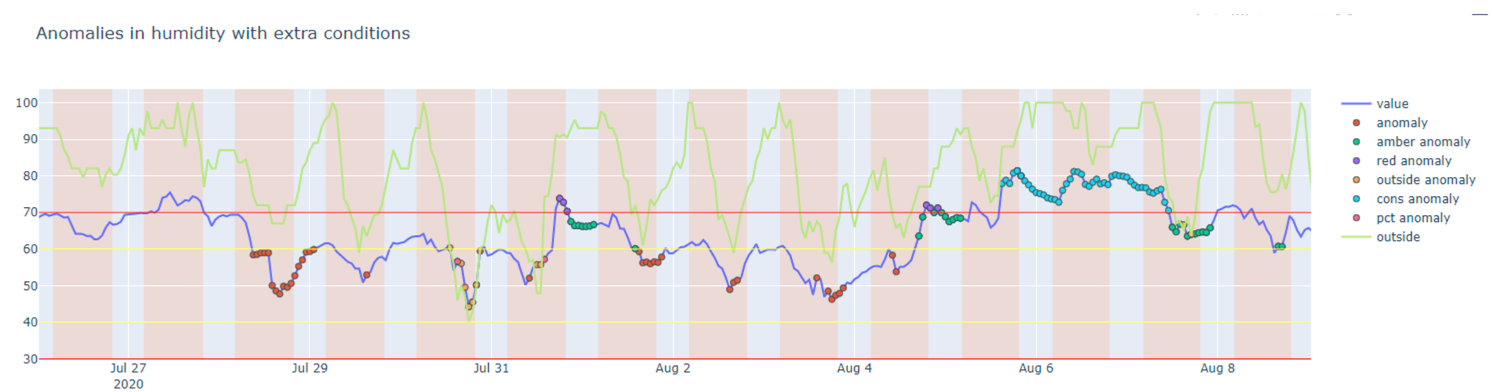


Figure 65: Property 1 Humidity with all conditions

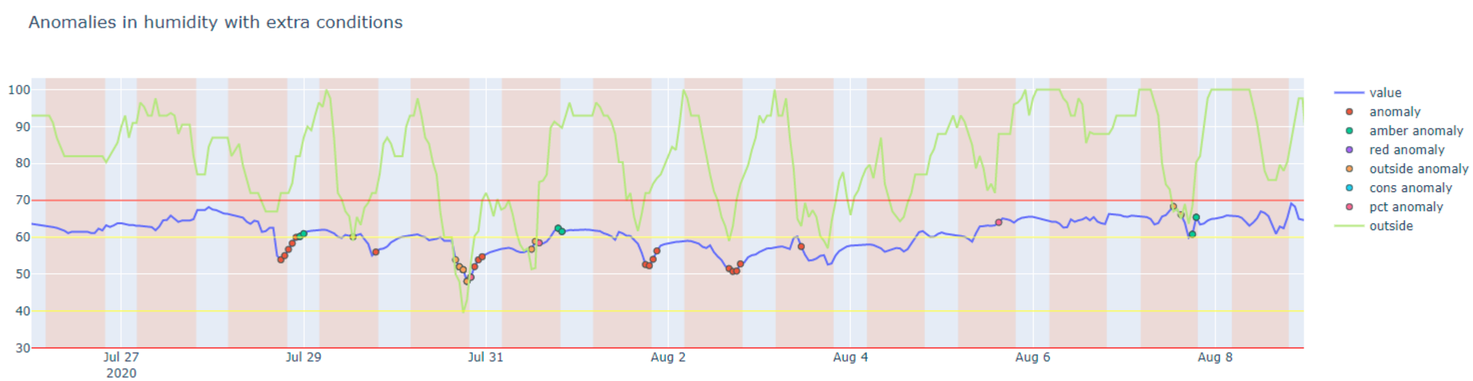


Figure 66: Property 2 Humidity with all conditions

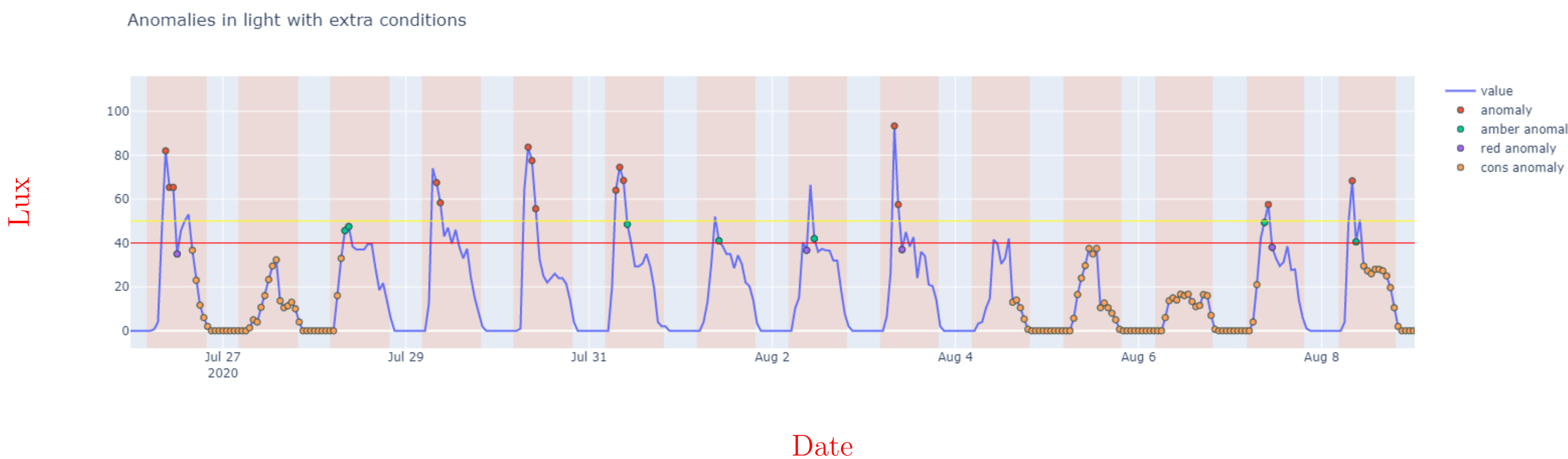


Figure 67: Property 1 Light with all conditions

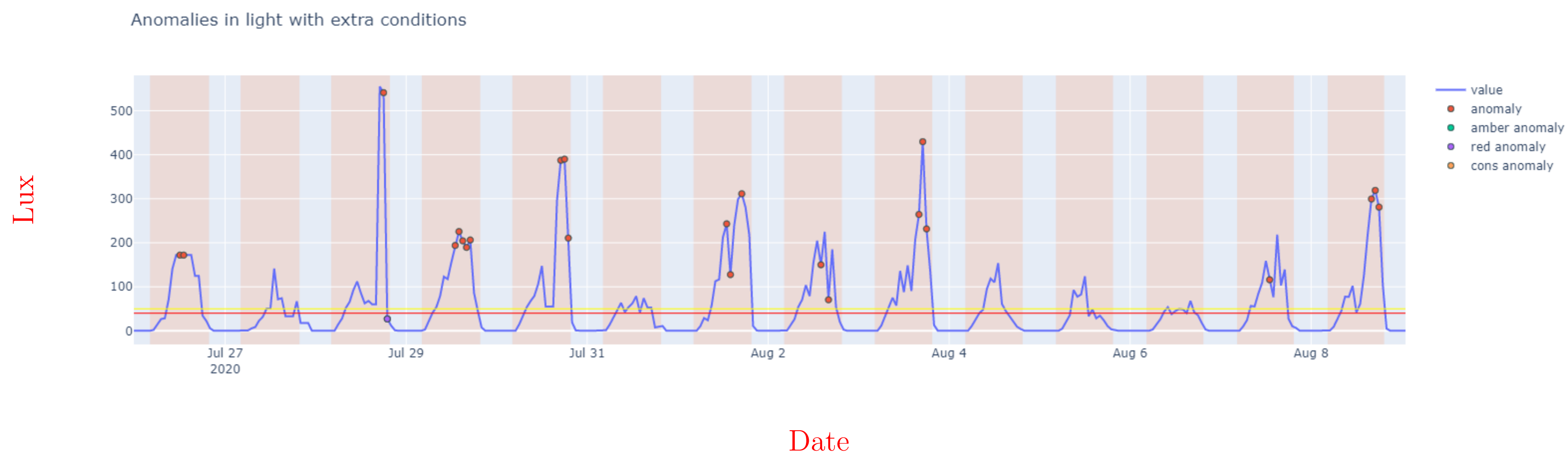


Figure 68: Property 2 Light with all conditions

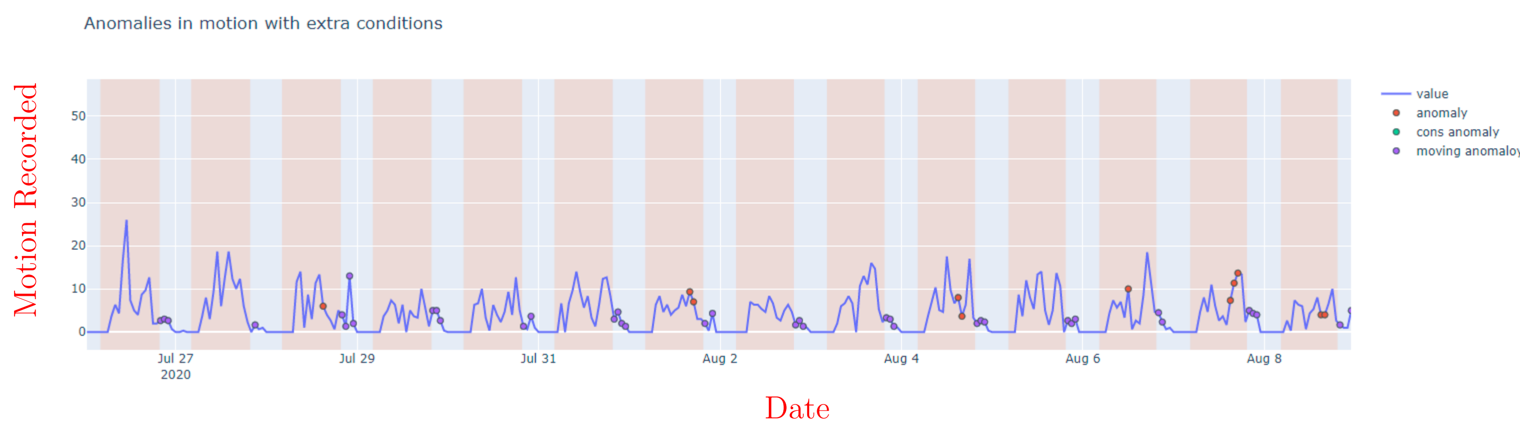


Figure 69: Property 1 Motion with all conditions

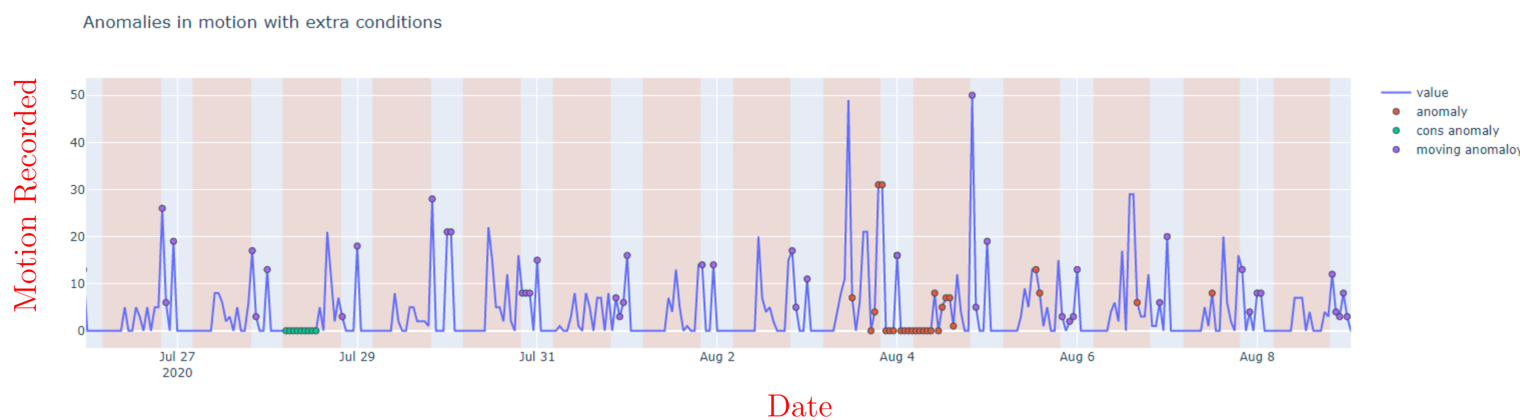


Figure 70: Property 2 Motion with all conditions