Interreg
North-West Europe
REAMIT
European Regional Development Fund

EUROPEAN UNION

WP T2 - Deliverable 8.1

Deployment of the integrated IoT, Big Data, analytics, Decision support technology

Improving Resources Efficiency of Agribusiness supply chains by Minimizing waste using Internet of Things sensors (REAMIT)

Del 8.1: Deployment of the integrated IoT / Big Data / analytics / Decision support technology

REAMIT: Improving Resource Efficiency of Agribusiness supply chains by Minimising waste using Big Data and Internet of Things sensors

2023

Interreg
EUROPEAN UNION
North-West Europe
REAMIT
European Regional Development Fund

Table of Contents

HMF *5*

Preliminaries **5**

Technology stack selected **6**

Analytics **7**

Pre-processing 7

Analysis 7

Interpretation **8**

Deployment **9**

Yumchop *10*

Preliminaries **10**

Analytics **10**

Pre-processing 10

Analysis 10

Interpretation **12**

Biogros *15*

Preliminaries **15**

Technology stack selected **16**

Analytics **16**

Pre-processing 16

Analysis 16

Interpretation **18**

Deployment **19**

Musgrave *22*

Preliminaries **22**

Technology stack selected **23**

Analytics **23**

Pre-processing 23

Analysis 23

Interpretation **23**

Deployment **23**

WD Meats & Burns Farm Meats *24*

Preliminaries **24**

Technology stack selected **25**

Analytics **25**

Pre-processing 25

Analysis 25

Interpretation **26**

Burns Farm Meats **27**

Alerting logic **27**

Picnic* *30

Preliminaries **30**

Technology stack selected **31**

Analytics **31**

Pre-processing 31

Analysis 31

Interpretation **31**

Deployment **32**

Raman* *33

Preliminaries **33**

Technology stack selected **33**

Analytics **33**

Pre-processing 33

Fig. The necessary steps to pre-process Raman spectra 34

Analysis 34

Interpretation **36**

Deployment **36**

HMF

Preliminaries

The Human Milk Foundation (HMF) plays a critical role in providing rapid delivery of breast milk to neonatal wards in South England using motorcycles and cars. However, due to a lack of means of temperature monitoring, there is presently a risk of spoilage if their cold storage fails without their knowledge until the milk is delivered to the hospital. To help reduce the chances of undetected cold storage failures during transportation, the REAMIT team proposed deploying temperature and humidity IoT sensors with HMF allowing them to monitor the parameters of the storage conditions and receive alerts if problems arose during milk delivery. With the temperature tracking sensors planned for use at HMF, however, another opportunity was discovered using the collected data.

Through analysis of the external parameters and their impact on ambient temperature changes during milk transportation, it was identified that the collected data from the temperature tracking sensors could be combined with other relevant business data collected on each journey and utilised to develop a regression model. With the regression model, HMF would have a quantitative tool to estimate, based on the specific planned journey parameters, the maximum duration and distance for delivery journeys without compromising the quality or safety of the milk due to temperature threshold abuse. This predictive capability would allow Human Milk Bank (HMF) to optimise their delivery planning and potentially extend their reach to sites located further away. Overall, the model could help HMF avoid breaches in temperature thresholds that could result in milk wastage by offering a tool to safely plan logistics so that this doesn't occur. Ultimately, the proposed model could enhance the operational efficiency of the milk delivery process, reduce the risk of milk wastage, and contribute to the overall success of the HMF. The development of a User Interface (UI) application was proposed, enabling end users to interact with the trained regression model and generate predictions for maximum journey length. Detailed below is a list of the equipment installed, the accompanying labels provided by HMF for each journey, the outline of the proposed analysis, and the data dictionary of all the variables in the complete dataset.

Sensors (x 10)

- Digital Matter eagle data logger
  - Internally records acceleration and GPS and reports a "trip" indication if these both update within the space of 5 minutes.
- T9603 T/RH probe
- Binary sensor providing bag lid status (1/0)

Business provided labels

- Date, time, and length of journey
- Weight of milk on journey
- Size of bag used for journey
- External weather temperature
- Type of vehicle used (car or motorbike)

Proposed Analysis

- Produce a regression model to estimate the maximum journey length based on user provided variables about the details of the journey. The model should be accessible and interfaceable through a user-friendly UI by staff at HMF so they can make use of the predictions provided by the model.

Data Dictionary

Source	Column	Data Type	Description

Sensor	device_id	Int	Numeric device ID, unique to the sensor
Sensor	datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
Sensor	battery	Float	Battery level during sensor recording
Sensor	temperature1	Float	Temperature recording inside bag
Sensor	humidity	Float	Humidity recording inside bag
Sensor	bag_status	char(1)	Indication if the bag is open (0) or closed (1)
Sensor	device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.
Sensor	Trip	char(1)	Indication if the bike is being driven (1) or is stationary (0) when the recording was made.
HMF spreadsheet	device_id		Numeric device ID, unique to the sensor
HMF spreadsheet	Datetime		Date and time of journey
HMF spreadsheet	journey_length		Approximate journey length in minutes
HMF spreadsheet	milk_quantity		Milk quantity on delivery (litres or KG)
HMF spreadsheet	bag_size		Small, medium, or large
HMF spreadsheet	vehicle_type		Bike or car
HMF spreadsheet	external_temperature		External weather temperature in Harpenden at time of milk departure

Technology stack selected

The analysis and application were developed using R, a popular open-source programming language known for its extensive capabilities in statistical analysis and machine learning. R provides a wide range of packages and libraries specifically designed for statistical modelling, making it a suitable choice for building predictive models. In addition to its statistical capabilities, R also offers the Shiny package, which allows for the development of interactive web applications. Shiny enables the creation of user-friendly front-end interfaces that can be deployed as web-based applications, providing a convenient way to showcase and interact with the models and their results. By utilising R and its Shiny package, the analysis and application development process can benefit from the rich statistical functionality of R while also providing an intuitive and interactive user interface through a web application, allowing for a comprehensive solution that integrates statistical modelling, data visualization, and user interaction.

Analytics

Pre-processing

The first step in producing the regression model was to prepare a training dataset for building the model. For preparing the training dataset, the data gathered from 117 trips was first compiled and aggregated. Then the data was pre-processed, cleaned, and transformed to a form that was useful for model building. The pre-processed dataset was divided into a training dataset and a test dataset. The ratio chosen for this split was 1:4, meaning that 20% of the data was allocated for testing the model's performance.

Analysis

After preparing the data, various modelling approaches were employed to develop predictive models. Namely, these approaches were linear regression, decision trees, random forest, and neural networks. Each model was trained using the prepared dataset, and their predictive performance was evaluated using the test dataset and the performance was quantified with several statistical measures. These were R^2 and RMSE, and for the linear model, VIF, Durbin Watson Test, F score, and P value were also checked.

- R-squared (R2) is a statistical measure that represents the proportion of the variance in the dependent variable (outcome) that can be explained by the independent variables (predictors) in the model. It provides an indication of how well the model fits the data. Higher values of R2 indicate a better fit, with 1.0 indicating a perfect fit.
- Root Mean Squared Error (RMSE) measures the average magnitude of the residuals (prediction errors) of the model. It provides a measure of the model's predictive accuracy, with lower values indicating better predictive performance. RMSE is calculated by taking the square root of the average of the squared differences between the predicted and actual values.
- To assess multicollinearity, the Variance Inflation Factor (VIF) was calculated. VIF measures the extent to which the variance of the estimated regression coefficients is increased due to multicollinearity in the predictors. A high VIF value (normally >5) suggests high multicollinearity, which can affect the reliability of the model.
- The Durbin Watson Test was performed to check for the presence of autocorrelation in the model's residuals. This test examines whether there is a systematic pattern or correlation among the residuals, which can impact the model's accuracy.
- The F Score and P value characteristics were also evaluated. The F Score is a measure of overall significance of the model, indicating whether the model as a whole is statistically significant. The P value helps determine the significance of individual predictor variables, indicating whether they have a significant impact on the outcome.

By analysing these statistical measures for each model, it was possible to assess their performance and select the most suitable model for the prediction task. Listed below are the results from each of these models.

	R2	**RMSE (°C)**	**VIF**	**Durbin Watson**	**F-score**	**p-value**
Linear regression	0.7396	0.7242	Time: 1.376079 Temp.		550.5	< 2.2e-16

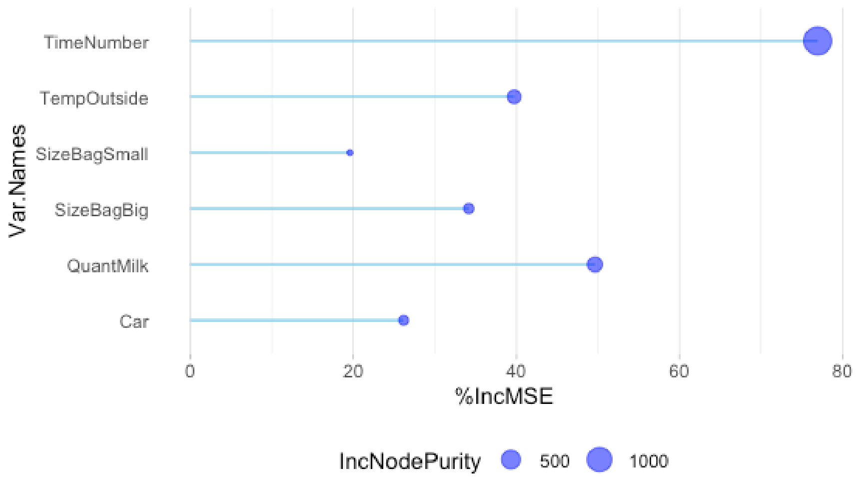
			Outside 1.549767 Quantity of milk 1.753195 Car: 1.377830 Size of bag (big): 1.493121 Size of bag (small): 1.446634
Decision trees	0.8721	0.5332	Not applicable
Random forest	0.9673	0.3585	Not applicable/ Non-linear
Neural networks	0.9514	0.3099	Not applicable/ Non-linear

With the given data, in our analysis we found linear regression and random forest algorithm based models provided the best accuracies.

In the codebase, createModel.R is the script which builds the model and model.rda is the resulting built model object.

Based on the code script in createModel.R, linear regression or Random forest, either of the models could be interchangeably used.

Interpretation



- Include visualisations
- Have we produced an analytical method which could be used to reduce food waste? Have we answered the open question posed by the pilot company?

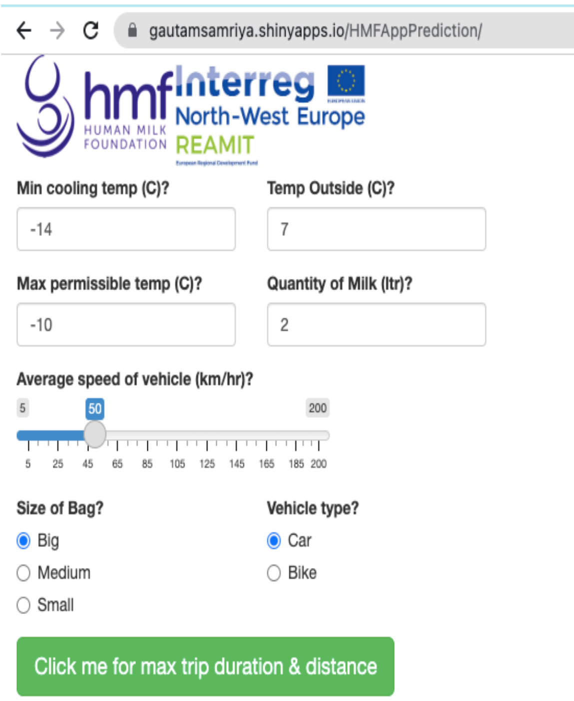
Deployment

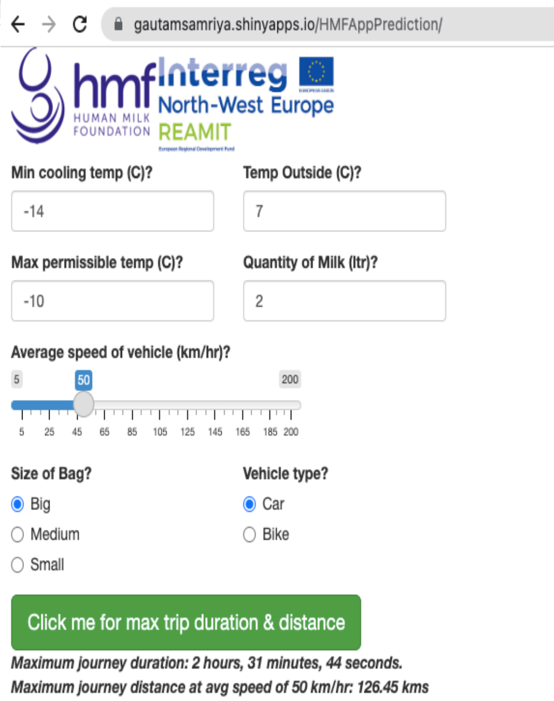
For simplicity, the application is designed in such a way that it provides a simple user interface to the user, who could be a delivery personnel or any other stakeholder. The UI, on taking inputs from the user, interacts with machine learning model in the backend, the details of which are hidden from the user. The user needs to input the value of external parameters for the journey, and then press the button. Upon button click, the application will send the user entered data on UI to the machine learning model script deployed in the backend as an input. The model will take these input values, perform the prediction and send the prediction result back to the user screen, which will display the resulted prediction results to the user. Thus for end user, the application looks simple and sleek and it hides the complexity of backend from the user. By hiding the model and not exposing the trip data, the application also ensures security.

The User Interface has been deployed using R Shiny, a web application framework for R programming language. The codebase includes an app.R script, which is responsible for creating the user interface and establishing the connection with the trained prediction model in the backend.

The application is accessible from the following URL:

https://gautamsamriya.shinyapps.io/HMFAppPrediction/





Yumchop

Preliminaries

Yumchop specialises in producing African flavoured frozen ready meals. They provide flavoursome and authentic food from around the world with an African twist that is frozen to retain its goodness and freshness and minimise waste. Yumchop is using locally sourced raw materials to prepare their ready-meal products. Most of the ingredients are supplied by local vendors, located locally from the production plant of Yumchop in Towcester, Northamptonshire, UK. Their tasty meals are distributed at institutions such as universities or hospitals through self-service automated vending machines. These unattended retailing kiosks have been fitted with an integrated microwave oven which enables them to warm the food upon purchase. However, Yumchop also delivers food to customers' homes through direct purchase at their website, enabling one-off purchases and monthly subscriptions that customers can customise to receive food at their preferred intervals. Moreover, they also supply directly to retailers and large organisations.

Equipment

- Sensors (x 10)
  - ELT Internal Antenna sensors (Elsys, Sweden)
  - DS18B20 temperature probes added to each ELT sensor.
  - 1 sensor was deployed per fridges/freezer
  - 10 sensors in total were used.

- Tektelic Kona Micro IoT Gateway (Tektelic, Canada)

Business provided labels

- None

Data Dictionary

Column	Data Type	Description
device_id	int	Numeric device ID, unique to the sensor
datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
battery	float	Battery level during sensor recording
temperature	float	Temperature recording inside chamber
device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.

Analytics

Pre-processing

- Data cleaning and transformation

Analysis

In this pilot test, no specific information was provided regarding the specific analytical applications that would be beneficial for the data. However, despite identifying anomalies, none of them yielded significant interpretability. Therefore, our focus shifted towards designing tools to effectively describe the data.

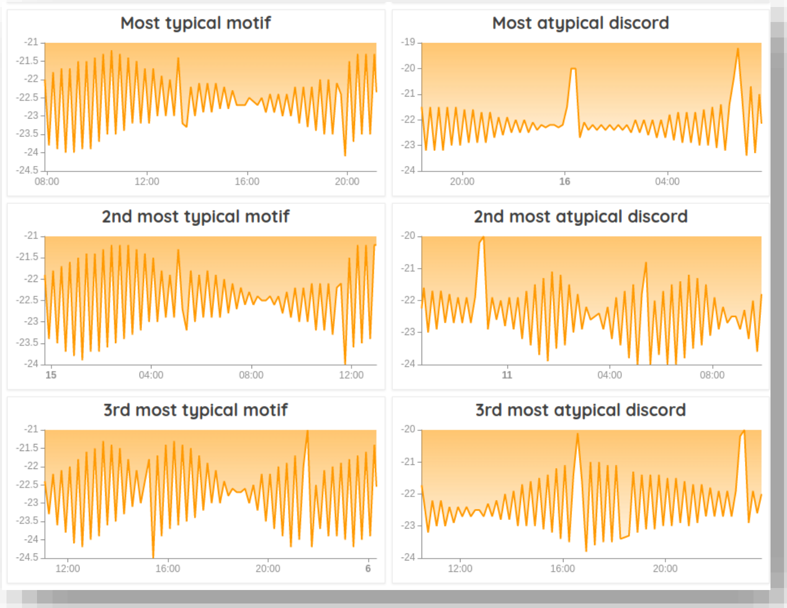
In the following figure, we have computed various features to provide a comprehensive description of the data. These features include the mean, standard deviation, minimum, maximum, median time between each record, potential seasonal cycles and associated weights.

In [4]:
```
%%warpscript -v
$data @reamit/analysis/synthesis
```

In [5]:
```
d = stack.pop()
df = pd.DataFrame.from_dict(d, orient='index').transpose()
df
```

Out[5]:

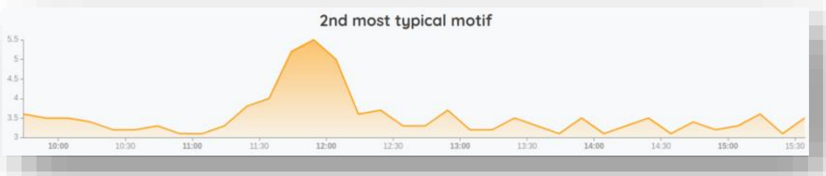
	classname	device_name	device_id	size	modes	mean	std	min	max	step_(s)	seasonality_1	val_1	seasonality_2	val_2
0	YUMCHOP.temperature	Vending Machine (066E)	A81758FFFE06066E	2003	48	-22.5	0.93	-24.6	-19.2	600.02	PT20M	0.69	PT19H10M	0.47
1	YUMCHOP.temperature	Zone D Fridge (066D)	A81758FFFE06066D	2009	79	-4.1	1.38	-5.5	11.0	599.99	PT1H30M	0.55	PT23H50M	0.32
2	YUMCHOP.temperature	Zone E fridge (0670)	A81758FFFE060670	2005	49	3.0	0.91	1.3	7.1	599.98	PT6H10M	0.4	PT6H10M	0.4
3	YUMCHOP.temperature	Zone B freezer 2 (0659)	A81758FFFE060659	2008	68	-30.0	1.51	-32.2	-25.5	600.03	PT146H5M	0.41	None	None
4	YUMCHOP.temperature	Zone E Freezer (065C)	A81758FFFE06065C	1996	60	-31.6	1.07	-32.8	-25.5	600.04	PT165H45M	0.36	PT120H25M	-0.01
5	YUMCHOP.temperature	Container cold room (0658)	A81758FFFE060658	1989	100	6.7	2.01	2.3	12.2	599.98	PT144H55M	0.25	PT94H50M	0.17
6	YUMCHOP.temperature	Zone D Coldroom Fridge (065B)	A81758FFFE06065B	2008	81	3.0	1.45	1.3	12.2	599.98	PT21H35M	0.46	PT21H35M	0.46
7	YUMCHOP.temperature	Zone D cold room freezer (05F7)	A81758FFFE0605F7	1976	70	-19.0	1.24	-22.0	-4.9	600.01	PT4H	0.45	PT4H	0.45
8	YUMCHOP.temperature	Zone B Freezer 1 (066F)	A81758FFFE06066F	2012	31	-23.6	0.66	-25.3	-22.0	600.02	PT30M	0.97	PT14H	0.46
9	YUMCHOP.temperature	A81758FFFE06065A	A81758FFFE06065A	2003	151	16.2	4.14	12.0	27.2	599.97	PT24H10M	0.2	PT126H30M	0.13
10	YUMCHOP.temperature	Zone B Standing freezer (066C)	A81758FFFE06066C	2011	105	-22.5	1.73	-26.1	-8.2	600.02	PT8H15M	0.57	PT8H15M	0.57



Next, we proceeded to explore patterns within the data using the matrix profile algorithm. This approach allowed us to identify recurring patterns as well as atypical occurrences. A resulting output of this analysis is depicted by the side screenshot, providing a visual representation of frequently occurring patterns and anomalies within the data.

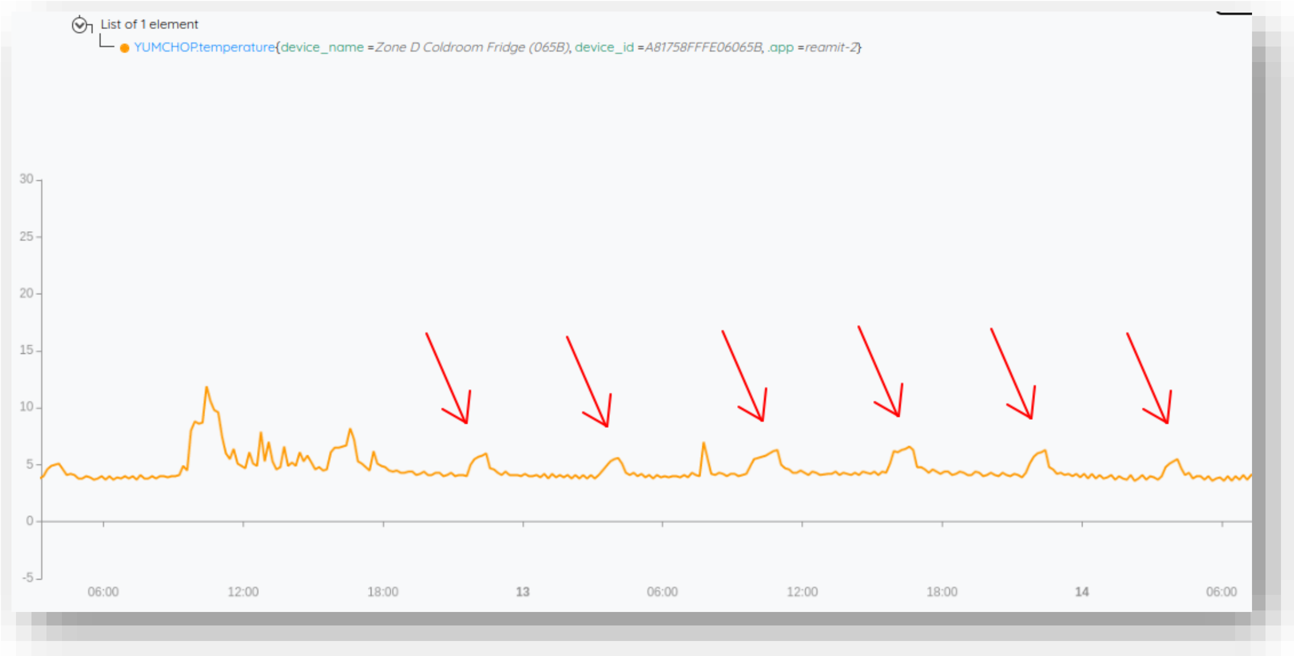
These representations effectively capture the data patterns for each sensor and highlight any deviations from those patterns, hence being useful for exploratory data analysis.

Interpretation



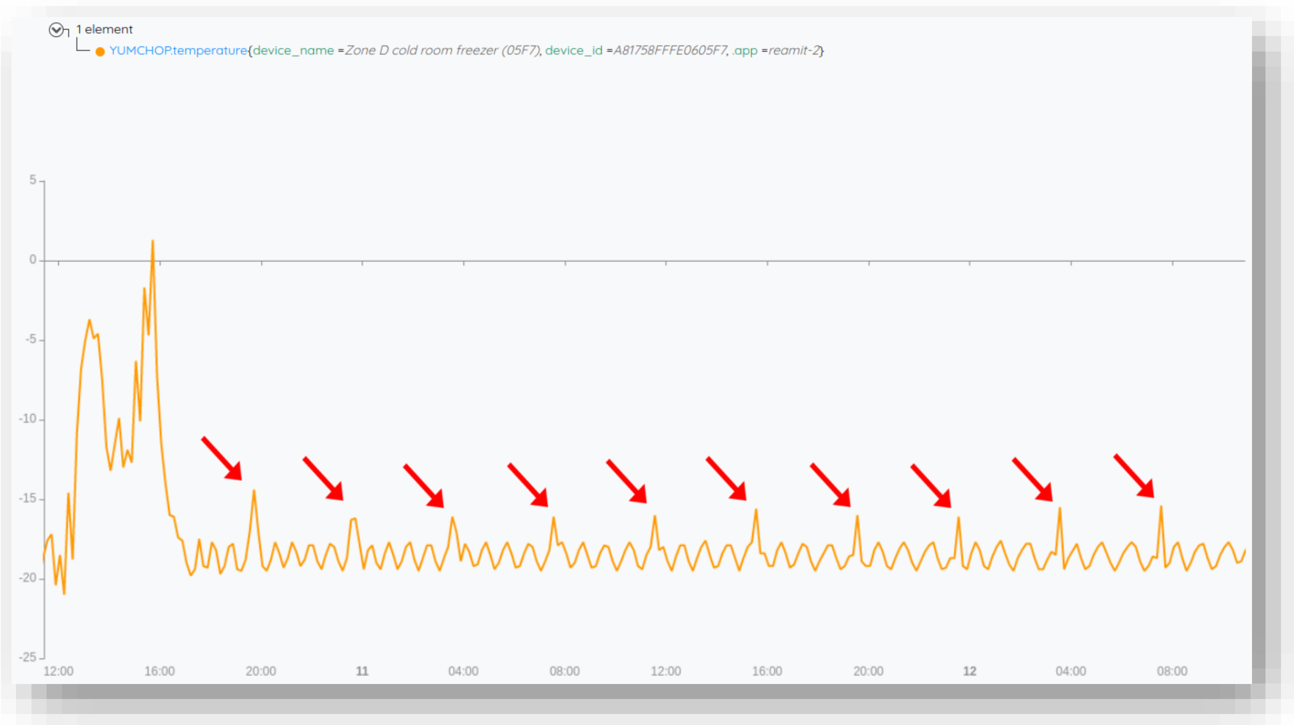
For sensors placed in refrigerators, the motif that typically appears as the second pattern in the matrix profile is likely indicative of a defrost operation.

In the following figure, we can identify instances of this motif:



We can notice a consistent occurrence pattern of this motif, indicating that the defrost operation typically takes place every 6 hours under normal circumstances. In essence, this enables us to predict the timing of the next defrost operation unless any external event disrupts this pattern.

Likewise, when considering sensors placed in cold room freezers, we can observe consistent motifs that likely reflect the effects of defrost operations on temperature, as shown in the following figure.



In this particular context, these motifs tend to occur approximately every 4 hours.

Additionally, it is worth noting that occasionally, atypical discordant patterns like the one observed at the beginning can be found in the data. These patterns are likely a result of significant movements within the cold room and instances where the door remains open for a long time.

Biogros

Preliminaries

- This is where you should describe the company, the problems they are facing, the data that has been collected (data dictionary), and the proposed analytics. The systems requirement specification should provide some of this information.

Biogros is a well-established company that has been supplying high-quality organic food to customers in Luxembourg for over 25 years. They are known for their commitment to customer satisfaction, regularly adapting their product offerings to meet the evolving needs of their clientele. Biogros stocks a wide range of organic products, including well-known brands like Naturata, Rapunzel, and Lebensbaum, as well as products from smaller, lesser-known producers. Through their close collaboration with the Bio-Bauere-Genossenschaft Lëtzebuerg (BIOG), a cooperative of organic farmers, they also offer a diverse selection of regional organic products under the BIOG brand. With an extensive inventory of over 3,500 items in categories such as fruits and vegetables, dry goods, and dairy produce, customers can find a wide variety of organic options.

Biogros has established a commendable track record in sustainable transportation practices and minimizing waste packaging. To further their commitment to reducing food waste, they have partnered with REAMIT to explore potential solutions. As part of this collaboration, Biogros has installed humidity and temperature sensors in key areas such as their warehouse, cold stores, and refrigerated trucks. These sensors enable continuous monitoring and generate valuable data that the REAMIT team can analyze and leverage in their efforts to find innovative applications aimed at reducing food waste.

Following an in-depth data analysis, the REAMIT analytics team has discovered two promising avenues. One of the contributors, SenX, has developed an innovative anomaly scoring system using temperature sensor data. This system employs logical ranges and multiple temperature thresholds, drawing inspiration from the basic artificial neuron model by McCulloch & Pitt. With the inclusion of a multi-scaling layer that segregates information within each range, the score proves useful for anomaly detection, individual sensor forecasting and quality assessment. For example, by leveraging this approach, we identified stationary temperature sensors in cold storage facilities that exhibit seasonal forecast patterns. Biogros has provided a physical explanation for these patterns. We also detected transportation periods where the quality of the refrigeration was not in the up to the desired standards.

Furthermore, Biogros is actively exploring opportunities to incorporate human feedback into their transportation operations. This feedback, specifically related to quality complaints regarding food delivery, can be instrumented to collectively refine the AI model based on each individual sensor's anomaly score neuron model. This would prove useful for them for root cause analysis across transportation routes and storage facilities.

Data Dictionary

Column	Data Type	Description
device_id	Int	Numeric device ID, unique to the sensor
datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
Battery	Float	Battery level during sensor recording
temperature	Float	Temperature recording inside truck or cold store
Humidity	Float	Humidity recording inside truck or cold store
device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.

Technology stack selected

- Data is pulled from Whysor API and ingested into a Warp 10 platform
- Anomaly score is implemented using warpscript
- Temperature thresholds are stored in time series metadata and can be updated with an HTTP call
- Forecasting and anomaly detection is implemented in warpscript
- Visualisation is provided by Discovery and WarpStudio

Analytics

Pre-processing

- Date is parsed and formatted as Unix timestamp in microseconds.
- Time series model is applied. One time series per sensor
- A mask is applied on time series related to sensor in moving vehicle to only keep data points when the sensor is moving
- The series are split into sub series by detecting quiet periods of 4 hours with no data
- Synchronization is applied by affecting the last recorded value to every tick every 5 minute
- Linear interpolation fills missing values

Analysis

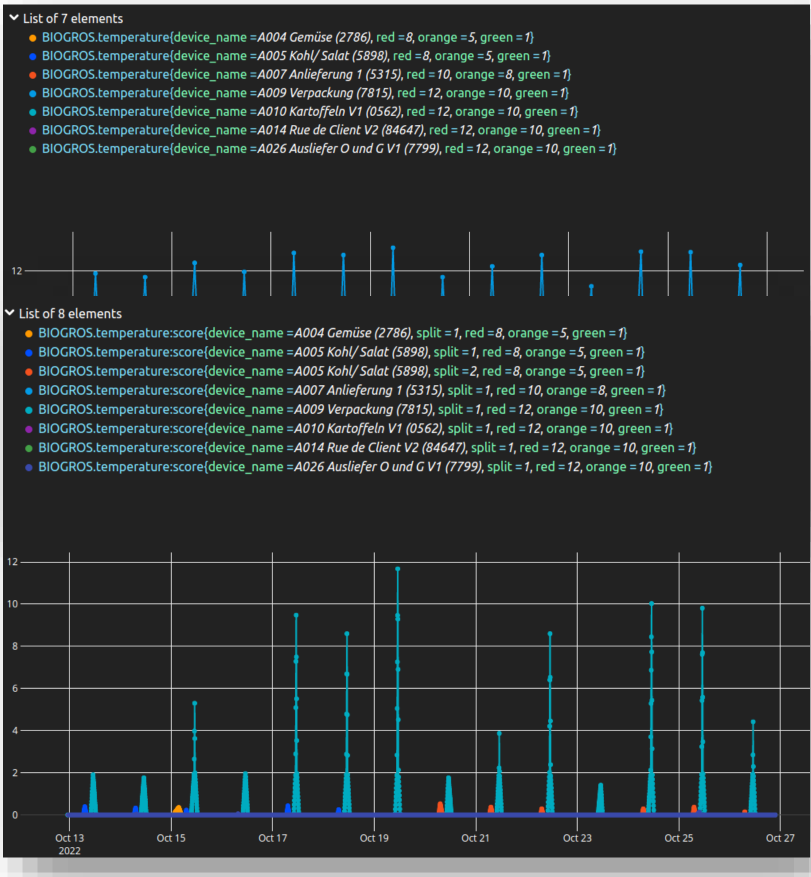
For each sensor, a set of two thresholds is provided. If the first one is crossed (orange threshold), then the cooling condition is not ideal but okay. If the second one is crossed (red threshold), then the conditions are bad. The formula of the anomaly score weighs in the temperature surplus when crossing these thresholds:

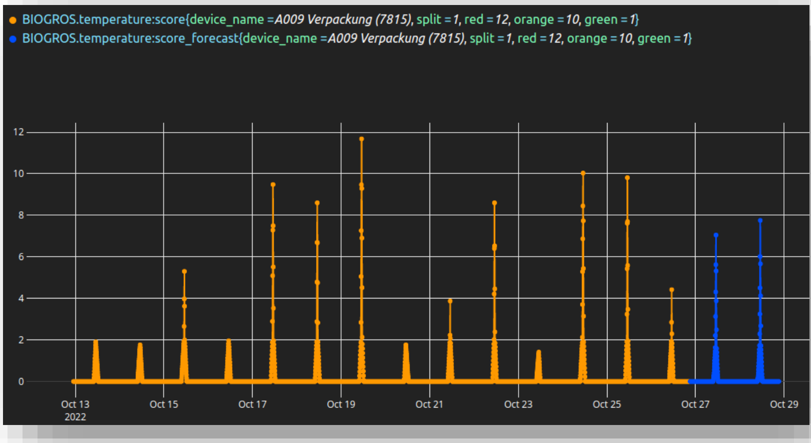
$$S(t) = \max\left(c_o \cdot reLU(t - t_o)\,,\; c_r \cdot reLU(t - t_r)\right)$$

where S is the anomaly score, t is the sensor's temperature reading, t_o is the orange threshold, t_r is the red threshold, and c_o, c_r are the associated penalty weights.

As a matter of fact, this formula can be interpreted as a basic two-neurons neural network, preceded by a multi-scaling operation and using max-pooling.

c_o and c_r are initialized at 1 and 10. Using consumer feedback, these values can be improved.

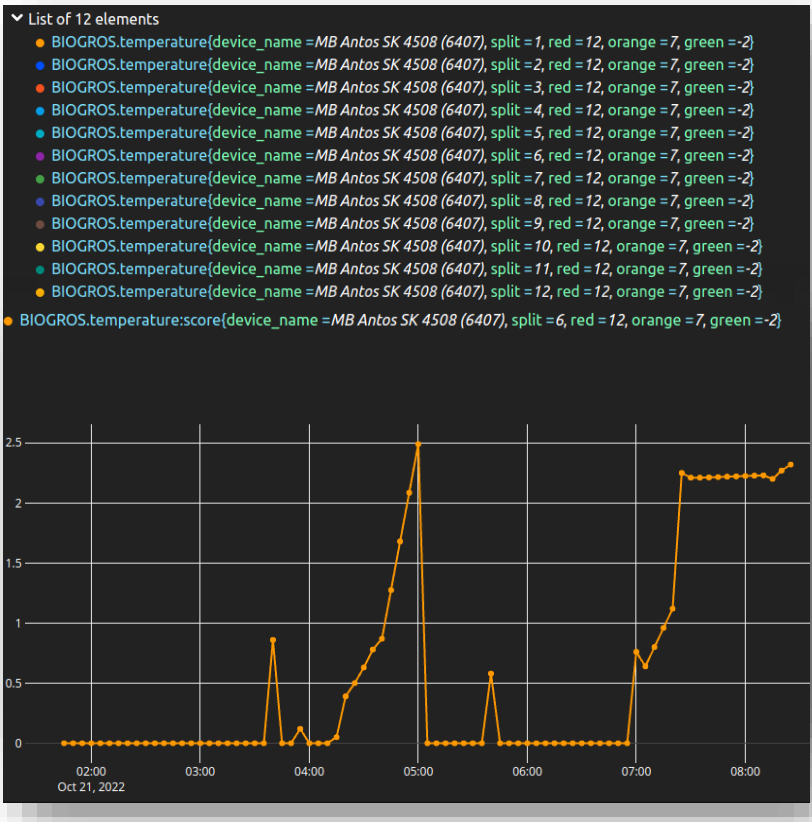




The first screenshot illustrates the plotting of data captured by temperature sensors in selected warehouses over a specific time period.

Afterward, the anomaly score is computed for each temperature time series by utilizing the logical ranges provided in the metadata of the time series. These scores effectively curate the time series data, enhancing the applicability of machine learning models.

An example of this is depicted in the provided screenshot, showcasing the plotting of the results generated by a forecasting model. Notably, this model demonstrates a clear seasonal forecasting pattern.



In this screenshot, an example of data from refrigerated trucks is plotted, specifically after applying a mask to retain only the data points corresponding to trips.

Presented here is an example of the anomaly score computed for a trip. A positive score indicates that the temperature exceeded the recommended threshold for transportation.

Interpretation

The analysis has revealed seasonal forecast patterns in the anomaly score for certain static sensors, indicating cyclical instances where the temperature exceeded the recommended threshold. These results were communicated to Biogros.

For moving sensors, a positive score suggests an inefficiency in freezing or cooling capabilities. As an illustration in the provided screenshot, the mean anomaly scores depict the quality of chilling conditions during product transportation. A score close to 0 indicates optimal chilling, meaning that the products were transported under suitable cooling conditions. If the score exceeds 1, it suggests that the chilling was acceptable but not ideal, indicating a need to slightly increase the cooling power. However, if the score surpasses 10, it usually indicates a lack of proper cooling throughout the transportation process.

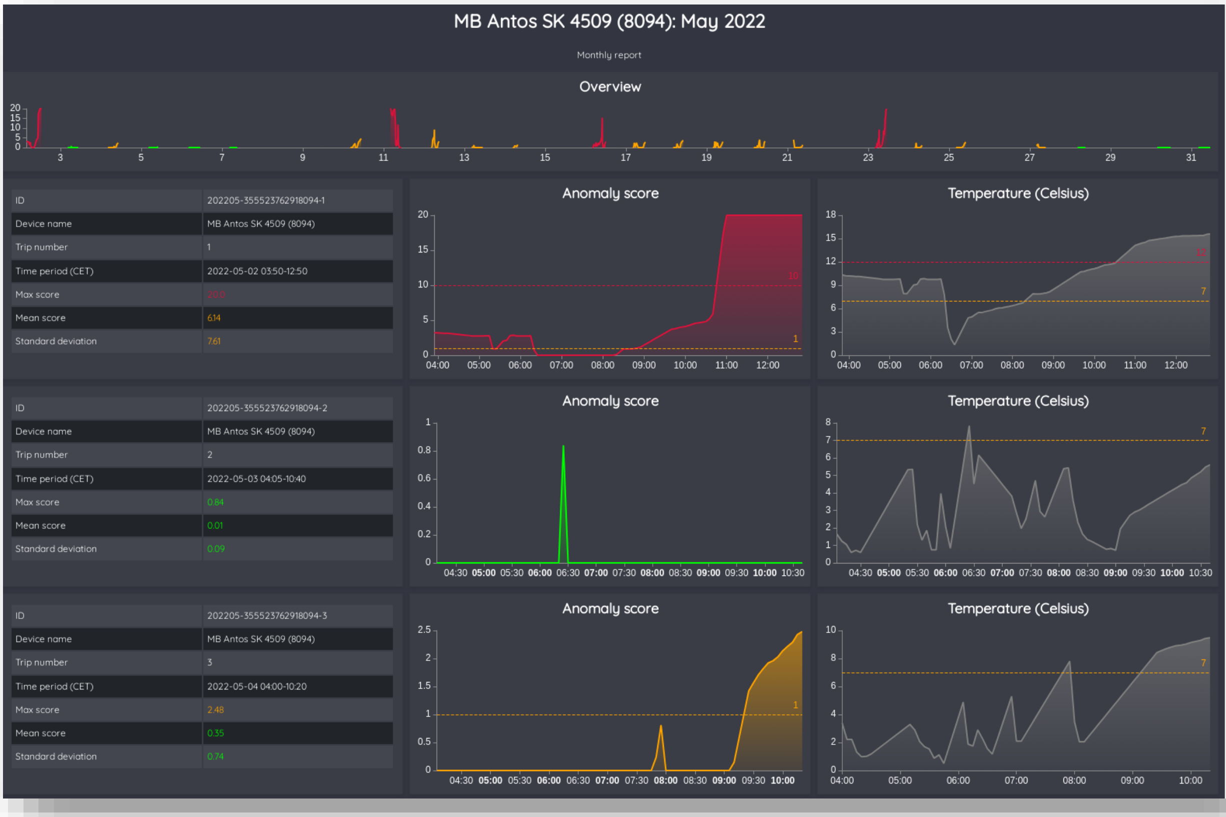
Mean anomaly scores

device_name	trip 1	trip 2	trip 3	trip 4	trip 5	trip 6	trip 7	trip 8	trip 9	trip 10	trip 11	trip 12	trip 13	trip 14	trip 15	trip 16	trip 17
MAN KN 8715 (2424)	0.4	0.77															
MAN ZE 0001 (6084)	0.05	1.54	0.74	4.78	2.46	13.86	12.9	75.07	2.68	3.2	0.47	2.36	2.7	2.07	4.0	1.9	0.29
MB Antos SK 4508 (6407)	55.46	0.39	0.18	1.7	5.03	0.56	15.13	0.59	0.13	0.59	0.03	0.98					
MB Antos SK 4509 (8094)	0.01	0.51	0.52	0.15	18.02	45.08	55.73	0.6	0.01	0.12	0.52						
MB Ateco SK 4463 (6480)	0.01	1.16	0.45	1.92	13.84	0.73	1.58	1.88	0.14	2.9	2.07						
MB YH 5823 (5995)	2.17	1.84	1.92	0.72	0.0	0.7	0.02	0.14	0.49								
MB YH 5868 (3274)	0.0	0.1	1.52	0.87	0.53	0.15	12.57	0.12	0.01								
Sprinter YH 5827 (8681)	0.55	0.84	4.1	1.5	2.5	0.76	1.85	3.77	0.39	11.19							

Deployment

Using Warp10 discovery, we have successfully deployed 126 reporting dashboards. These dashboards consist of two types of reports. The first type is a comprehensive monthly quality report that provides detailed information about the transportation of a specific truck. The second type is a monthly overview report that gathers anomaly scores for all trips that took place within the month.

Below are examples of these interactive reports, which allow users to interact with the graphs using the mouse pointer.


MB Antos SK 4509 (8094): May 2022
Monthly report
Overview
ID
202205-355523762918094-1
Device name
MB Antos SK 4509 (8094)
Trip number
1
Time period (CET)
2022-05-02 03:50-12:50
Max score
Mean score
6.14
Standard deviation
7.61
Anomaly score
Temperature (Celsius)
ID
202205-355523762918094-2
Device name
MB Antos SK 4509 (8094)
Trip number
2
Time period (CET)
2022-05-03 04:05-10:40
Max score
0.84
Mean score
0.01
Standard deviation
0.09
Anomaly score
Temperature (Celsius)
ID
202205-355523762918094-3
Device name
MB Antos SK 4509 (8094)
Trip number
3
Time period (CET)
2022-05-04 04:00-10:20
Max score
2.48
Mean score
0.35
Standard deviation
0.74
Anomaly score
Temperature (Celsius)


Overview: April 2022
Monthly anomaly scores
MAN KN 8715 (2424)
MAN ZE 0001 (6084)
MB Antos SK 4508 (6407)
MB Antos SK 4509 (8094)
MB Ateco SK 4463 (6480)
MB YH 5823 (5995)
MB YH 5868 (3274)

Musgrave

Preliminaries

At Musgrave, last mile refrigerated delivery vehicles perform deliveries within the greater Belfast area. Presently, Musgrave have no means of temperature monitoring, meaning if a refrigeration unit has malfunctioned, they are unaware until the van makes its next stop. This is often too late and could result in hundreds of kilograms of spoiled stock. Instead, they wish to have an automated instant alerting system so that they know if the temperature in the rear of the van reaches above a defined threshold. Detailed below is a list of the equipment installed, the outline of the proposed analysis, and the data dictionary of all the variables in the complete dataset.

Sensors (x 5)

- Digital Matter eagle data logger
  - Internally records acceleration and GPS and reports a "trip" indication if these both update within the space of 5 minutes.
- DS18B20 Temperature probe
- T9603 T/RH probe

Business provided variables

- None

Data Dictionary

Column	Data Type	Description
device_id	Int	Numeric device ID, unique to the sensor
datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
battery	Float	Battery level during sensor recording
temperature1	Float	Temperature 1 recording, in freezer of van
temperature2	Float	Temperature 2 recording, in fridge of van
humidity	Float	Humidity recording, in freezer of van
device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.
trip	char(1)	Indication if the van is being driven (1) or is stationary (0) when the recording was made.

- This is where you should describe the company, the problems they are facing, the data that has been collected (data dictionary), and the proposed analytics. The systems requirement specification should provide some of this information.

Technology stack selected

Matlab 2021b was chosen as the software for analysing the last mile delivery data due to its popularity and strong reputation in the field of data analysis. Matlab is widely recognized for its advanced statistical analysis toolkits, which provide a comprehensive set of functions and algorithms for conducting various statistical analyses. Furthermore, Matlab is renowned for its advanced data visualization capabilities. It offers a wide range of plotting and graphing functions that allow researchers to visually explore and present the cooling profile data in a clear and meaningful way. The interactive nature of Matlab's data visualization tools makes it easy to customize plots, add annotations, and interact with the data, facilitating a deeper understanding of the underlying patterns and trends within the cooling profile data.

Analytics

Pre-processing

- Data cleaning and transformation

Analysis

- Application of ML / statistical techniques to identify trends, production of model, etc.
- Include visualisations

Interpretation

- Draw conclusions from the analysis
- Include visualisations
- Have we produced an analytical method which could be used to reduce food waste? Have we answered the open question posed by the pilot company?

Deployment

- If achieved, the pilot test lead can report on successful deployment of their model for online real time analytics (HMF, Raman)

WD Meats & Burns Farm Meats

Preliminaries

Located in Coleraine, Northern Ireland, WD Meats have been supplying quality assured beef and innovative beef products to retail customers throughout UK, Europe, Africa and Asia for over 35 years. WD Meats select only the best local beef that Northern Ireland has to offer. Then they process and deliver it with the utmost care and attention to the animals, and to the highest standards that their customers demand.

At WD Meats, there was an opportunity for optimising the dry-ageing process. This is a 28-day cycle used for premium cuts of beef to both improve the tenderness of the beef and enhance the flavour, achieved by maintaining temperature and humidity in a sealed, refrigerated room. The ideal parameters for this room are still being explored, and so the REAMIT team installed sensors to the room to help map the conditions during the current ageing process and thus help identify more "ideal" parameters to reduce the weight percentage lost during the process, while avoiding the dark-face phenomenon. "Dark facing" meat forms when too much moisture is drawn from the hindquarter, which must be cut off (or trimmed) prior to sale. This meat is classified as food waste. Detailed below is a list of the equipment installed, the accompanying labels provided by WD Meats and Burns Meats for trim loss, the outline of the proposed analysis, and the data dictionary of all the variables in the complete dataset.

Equipment

- Sensors (x 4)
  - Ursalink UC-11 Internal Antenna sensors (Ursalink, China)
  - The UC-11 consists of 2 built-in sensors: temperature, humidity
  - 4 sensors were deployed in the refrigeration chamber

- Multitech conduit mLinux IoT Gateway (Multitech, USA)

Business provided labels

- Hindquarter weights before and after dry ageing process

Proposed Analysis

- Carcass/hindquarter weight loss during the dry ageing process (loss in water content).
- Trim loss (crust/surface of the meat that needs to be trimmed off) after the dry ageing process.

Data Dictionary

Column	Data Type	Description
device_id	Int	Numeric device ID, unique to the sensor
datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
Battery	Float	Battery level during sensor recording
temperature	Float	Temperature recording inside chamber

Humidity	Float	Humidity recording inside chamber
device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.

- This is where you should describe the company, the problems they are facing, the data that has been collected (data dictionary), and the proposed analytics. The systems requirement specification should provide some of this information.

Technology stack selected

Matlab 2021b was chosen as the software for analysing the trim loss data due to its popularity and strong reputation in the field of data analysis. Matlab is widely recognized for its advanced statistical analysis toolkits, which provide a comprehensive set of functions and algorithms for conducting various statistical analyses. Furthermore, Matlab is renowned for its advanced data visualization capabilities. It offers a wide range of plotting and graphing functions that allow researchers to visually explore and present the trim loss data in a clear and meaningful way. The interactive nature of Matlab's data visualization tools makes it easy to customize plots, add annotations, and interact with the data, facilitating a deeper understanding of the underlying patterns and trends within the trim loss data.

Analytics

Pre-processing

- Data cleaning and transformation

Analysis

The first question to examine was the relationship between temperature in the dry-age chamber and the weight-loss during the dry-age process. Before and after weights were collected on one load of hindquarters, and the temperature during the dry-ageing process extracted. Figure X shows the temperature mapping over the 14-day dry-age process, which illustrates how the temperature is clearly colder at the front of the dry-age chamber closer to the refrigeration unit compared to at the rear of the chamber, furthest away from the dry-age chamber. Closest to the refrigerator, the mean temperature was -0.7954, std 0.3274 over the 14-day dry age period. In comparison, furthest from the refrigerator, the mean temperature was 0.0164, std 0.3459 over the same time period.

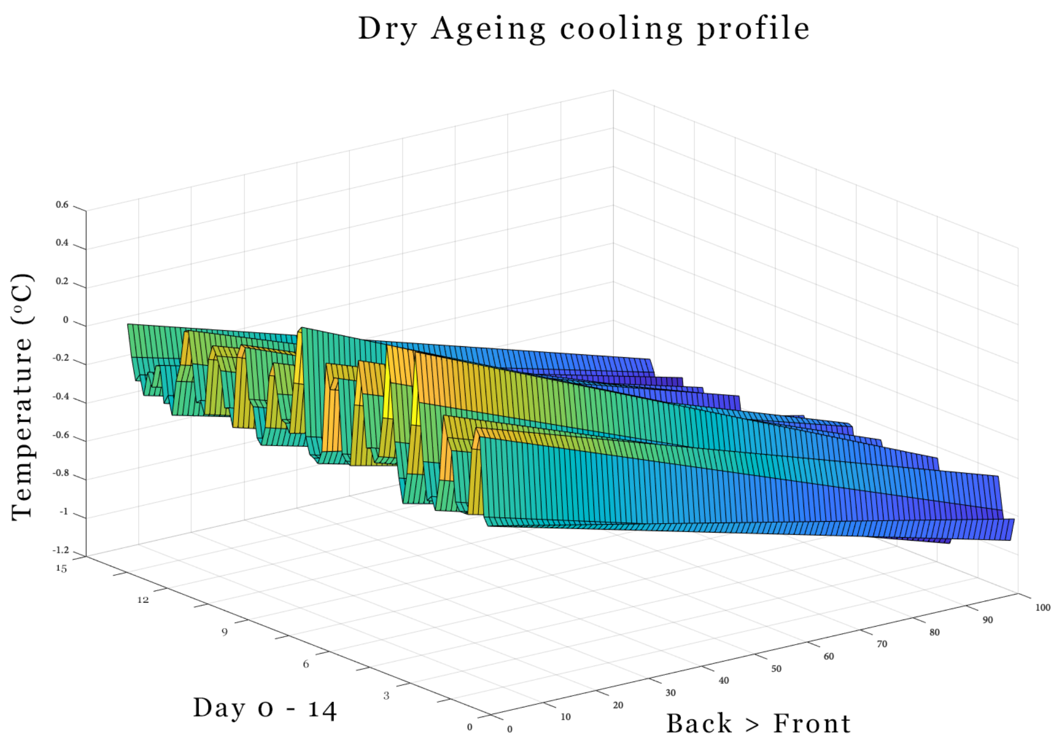


Figure 1: Dry ageing profile over a 14- day period.

The next step was to compare the weights of the hindquarters located in the front half of the chamber with those in the rear half. After the 14-day dry ageing was complete, the hindquarters in the front, cooler half of the chamber (closest to the fridge unit) had lost 3.87% total weight, while the hindquarters in the rear, warmer half had lost 4.33% total weight. An ANOVA (Analysis of Variance) test, a statistical test used to determine whether there are any significant differences between the means of two or more groups, was performed to check for statistical significance in weights between the hindquarters storage areas. The result of the ANOVA test was $P<0.05$, telling us that the differences in weight loss between the front and rear were statistically significant.

Interpretation

From this initial analysis, we have learned that the cooler the temperature in the dry-age chamber, the less weight loss occurs on the hindquarters. The next step of the analysis is examining how much trim loss is observed caused by dark facings on the hindquarters located at the front of the chamber in comparison to those at the rear. With this baseline dataset, parameter testing on the refrigeration settings can be performed to see if an ideal set of parameters can be located which minimises both the water weight loss and the trim loss caused by dark facings.

- Include visualisations
- Have we produced an analytical method which could be used to reduce food waste? Have we answered the open question posed by the pilot company?

Burns Farm Meats

Pilot testing of a similar nature was also conducted with the pilot test company Burns Farm Meats. Located in north Sligo, Ireland, Burns Farm Meat Ltd. is a small-sized, family-owned company with main activities including farming, operation of an abattoir, and processing of organic meats. As part of these activities, being firmly committed to animal welfare and providing meat of the highest quality, they run a dry-ageing process to deliver tender cut meat of their own locally raised, fed and cared for animals.

This pilot test mainly consisted in the monitoring of environmental parameters - temperature and humidity - in the two refrigeration chambers that they use to carry out dry ageing of beef. To this company, it was important to obtain a system capable of monitoring environmental parameters in their refrigeration chambers that would ultimately provide them with a higher understanding and the means to enhance the quality of meat and reduce its loss.

Equipment

- Sensors (x 10)
  - ELT-2 Internal Antenna sensors (Elsys, Sweden)
  - The ELT-2 consists of 4 built-in sensors including the two necessary parameters for the pilot test: temperature and humidity
  - 6 sensors were deployed in the company's larger refrigeration chamber
  - 4 sensors were deployed in the company's smaller refrigeration chamber

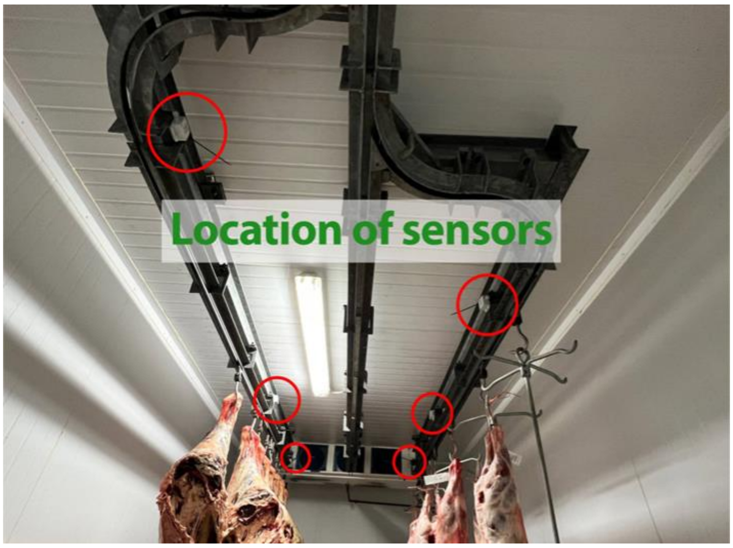


Figure xx: showing the location of 6 sensors in the company's larger refrigeration chamber.

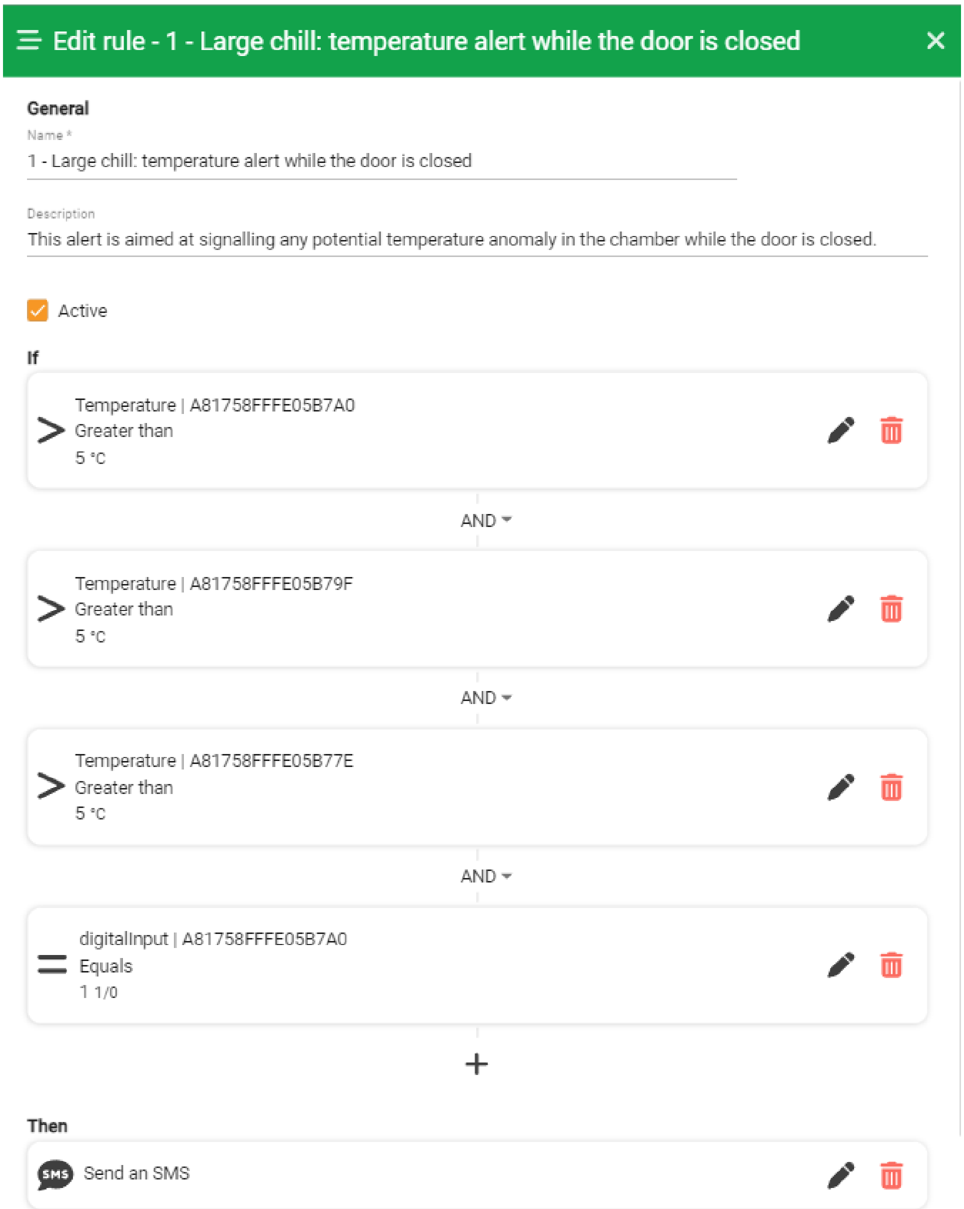
Alerting logic

One of the main REAMIT approaches to enhance efficiency and reduce food loss is the design and implementation of a suitable alerting system that can notify company representatives or staff members of potential anomalies in a timely manner. Following sensor installation, by continuously collecting data, the deployed sensors can identify potential issues or deviations that could lead to food spoilage by comparing each newly acquired environmental reading to a predefined safety threshold value. This way, real-time monitoring enables swift corrective actions, ensuring the quality and safety of food products.

In the context of this pilot test, and to best adapt to Burns Farm Meats operations and procedures, the alerting system had to be such that it would not send alerts repeatedly at times when they were aware that the chambers were not at their adequate temperature, e.g., they were carrying out cleaning, loading or unloading procedures and the door of the chamber would remain open, leading to an increase of temperature inside. In some instances, it could also be that the refrigerator had been turned off as the chamber was not in use – this mainly concerned the smaller chamber.

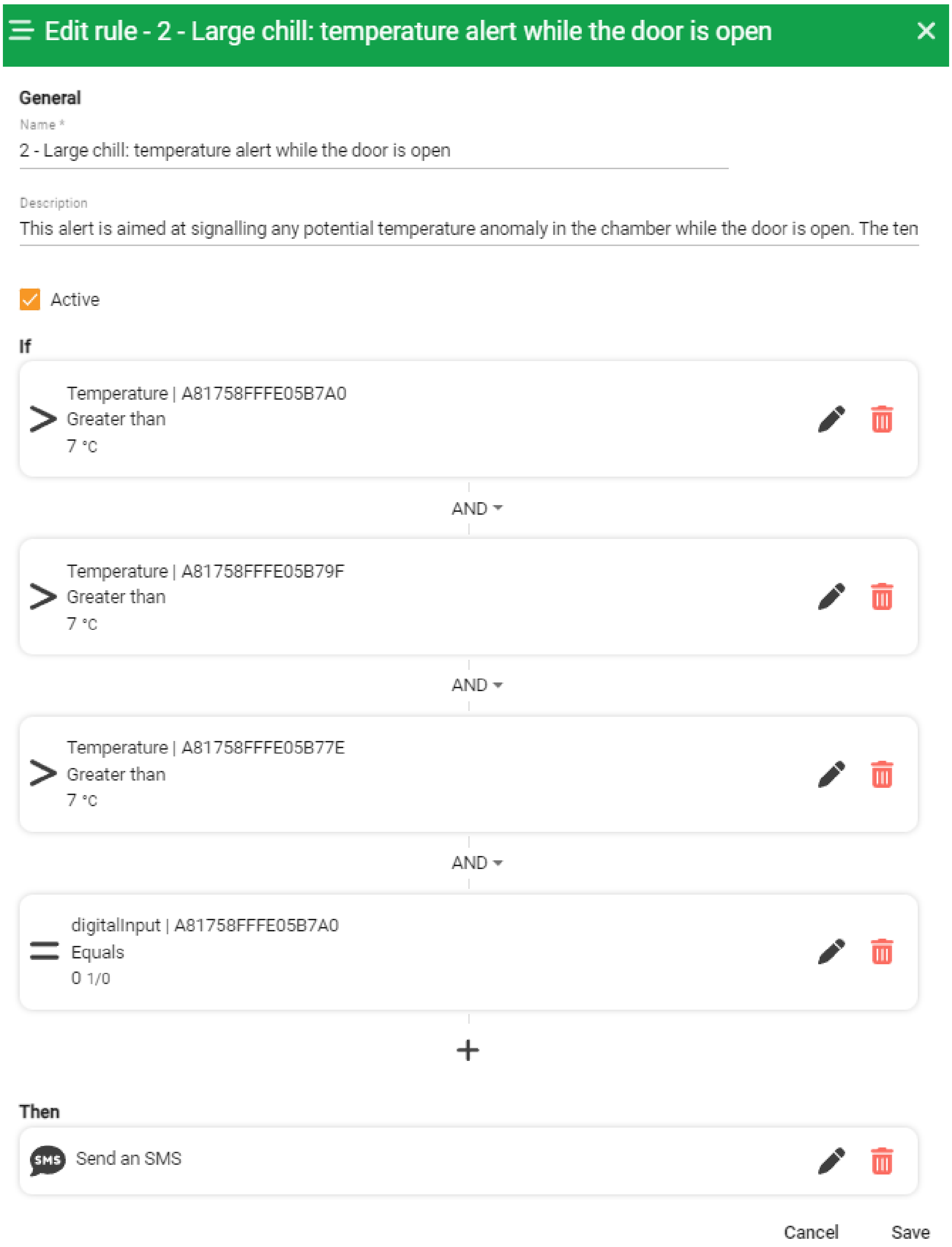
For this reason, the REAMIT team decided to install door contact switch sensors on the doors with the purpose of refining the logic of the alerting system.

The first of the two alerts that was pilot tested was designed to alert of anomalies while the door of the chamber was closed (figure xx).



In this case, the temperature threshold selected was 5°C and it had to be recorded consecutively 6 times by 3 different sensors. By implementing an "AND" logic, the team would avoid alerting if the selected sensor was malfunctioning. While that could also lead to not alerting

A second alert was also designed to signal potential issues while the door remained open.



Picnic

Preliminaries

Picnic is a pure online last-mile grocery delivery company. It offers its users to find and order grocery items through their mobile phone. The products that it markets are vegetables, fruits, meat, fish, sweets, snacks, drinks, dairy, bread, but also non-food items. The company has presence in the Netherlands, Germany and France. They strive to provide affordable and sustainable service to customers using cutting-edge technology, efficient planning, and a fleet of electric vehicles. A typical vehicle on grocery delivery would have many food boxes and each food box would have grocery for delivery to customers based on their orders. Chilled goods are stored in some boxes, goods that do not require cooling are stored in ambient temperature, while some other boxes store frozen food below prespecified temperature thresholds. To maintain the food quality inside delivery vans, it is necessary that temperature inside the cooling boxes is always maintained below a certain threshold level throughout the journey. The company uses ice packs to maintain the temperature inside the cooling boxes.

For efficient operation, the company needed an ability to continuously monitor the temperature in each cooling box, make sure that the temperature inside is within an accepted pre-defined threshold, and check whether the capability of the process of cooling is acceptable. The company also need optimization of ice-packs usage in the cooling process of its delivery vehicles.

Equipment

- Sensors (x 10)

EMS Internal sensors (Elsys, Sweden) (Humidity, Temperature, Acceleration)

- 1 sensor was deployed per crates, each box had 1 sensor installed.
LoRa coverage is provided by the netherlands KPN country-wide network.
- 10 sensors in total were used.

Business provided labels

Optimizing usage of cooling ice-packs
Ensuring high quality food supply(Reducing food wastage/Reducing customer complaints)

Business defined variables

- acceptable temperature range / temperature threshold

Proposed alerting logic

- If temperature exceeds threshold send alert (SMS/email)
- If there are too many shocks or disorientation of crates
- If cooling process is not in statistical control or lacking capability

Proposed Analysis

Statistical process control, and process capability analysis.
Alerting in case of temperature crossing acceptable threshold.
Acceleration along x,y,z axes based orientation analysis of boxes/crates.
Optimizing usage of cooling ice-packs.

Data Dictionary

Column	Data Type	Description

device_id	int	Numeric device ID, unique to the sensor
datetime_measure	ISO8601 date-time format, "yyyy-MM-dd'T'HH:mm:ss.SSS'Z'"	Datetime stamp of sensor recording
battery	float	Battery level during sensor recording
temperature	float	Temperature recording inside food crate/box
device_name	nvarchar(50)	Descriptive name of device. Should help identify where device is installed, but often left as device_id.
Motion_acceleration	int	The number of movement events during the measurement in a scale of 0 to 255
acceleration_x	Int	The forces on the X axis on this measurement in a scale of -127 to 127 (-2G to 2G, so 63 == 1G). This can be used to determine how the sensor has been placed, or if it's been handled roughly or moved
acceleration_y	Int	The forces on the Y axis on this measurement in a scale of -127 to 127 (-2G to 2G, so 63 == 1G).
acceleration_z	Int	The forces on the Z axis on this measurement in a scale of -127 to 127 (-2G to 2G, so 63 == 1G).
Humidity	Int	The humidity in %, in food crates/box.

- This is where you should describe the company, the problems they are facing, the data that has been collected (data dictionary), and the proposed analytics. The systems requirement specification should provide some of this information.

Technology stack selected

-Detail the tools and applications selected to be used for the analytics to provide useful insights for the pilot company

Analytics

Pre-processing

- Data cleaning and transformation

Analysis

- Application of ML / statistical techniques to identify trends, production of model, etc.
- Include visualisations

Interpretation

- Draw conclusions from the analysis
- Include visualisations
- Have we produced an analytical method which could be used to reduce food waste? Have we answered the open question posed by the pilot company?

Deployment

- If achieved, the pilot test lead can report on successful deployment of their model for online real time analytics (HMF, Raman)

Raman

Preliminaries

The pilot program at UoN focused on upholding food quality during storage and transportation. It placed particular emphasis on utilizing an efficient and rapid technique called Raman spectroscopy. To accomplish this, a comprehensive plan was formulated. The initial stage involved laboratory testing of the Raman sensor on food samples, analyzing the acquired data, and subsequently uploading it to the server. This phase, known as "lab development," was followed by evaluating the technology under conditions resembling those encountered in refrigerated trucks, referred to as the "transitioning phase." Lastly, the technology will undergo testing in real-world conditions, representing the third and final stage of the pilot program.

The final system developed as a result of passing through all above phases is composed of:

- Raman spectrometer (QE Pro, Ocean Optics) - used to acquire the raman spectra of a sample
- Fiber-optic probe (InPhotonics, RPB78)
- Laser 785 nm (Oxxius, LBX-785-HPE) - laser source
- Laser controller (Oxxius, LaserBoxx series, ControlBoxx for LBX models) - control the power of the laser
- 3-axis motorized platform (Standa, 8-0026) - allow to move the sample in the desired location (X, Y, Z)
- 3-axis controller (Standa, 8SMC4-USB-B8-B9)
- A lockable box (QUIPO Locker - dimensions: height 450 mm x width 450 mm x depth 450 mm) covered on the inside with black opaque adhesive film (Adhesive vinyl, ME310-61). The box is equipped with four adjustable legs. - To prevent light from entering and interfering with the Raman signal
- A Raspberry Pi Kit – Allow for the Raman system to be an IOT-complaint device.

The aforementioned system was tested on a variety of food products, including chicken, resulting in the collection of over 6000 Raman spectra. These spectra were pre-processed (baseline correction, smoothing and normalization) and then analyzed using multivariate data analysis techniques such as Principal Component Analysis (PCA) and Kruskal-Wallis (KW) test. To facilitate further analysis, an automated script was developed to extract the data and transmit it through a server. Additionally, this script is capable of pre-processing the Raman spectra, conducting analysis, and providing the aforementioned results within a timeframe of 2-5 minutes, depending on the data size.

Technology stack selected

- Training data was provided offline by University of Nantes and ingested into a Warp 10 platform
- Python was used to design a pipeline for data ingestion. Python was later used at the model deployment stage.

Analytics

Pre-processing

Raman spectra were analyzed using Opus software (Bruker optics GmBH, V 7.2, Germany) and MATLAB software (version R2019b, MathWorks Inc, Natick, MA, USA). Initially, the raw data spanning a spectral range of 200–4000 cm^{-1} was restricted to the range of 500–3000 cm^{-1}. Subsequently, any interference caused by cosmic spikes within the defined spectral range (500–3000 cm^{-1}) was manually removed prior to further data processing. To enhance the accuracy of the analysis, all spectra were subjected to baseline correction using an elastic concave method (with parameters of 64° and ten iterations), followed by smoothing using the Savitzky-Golay algorithm. Finally, the spectra were normalized using min–max normalization.

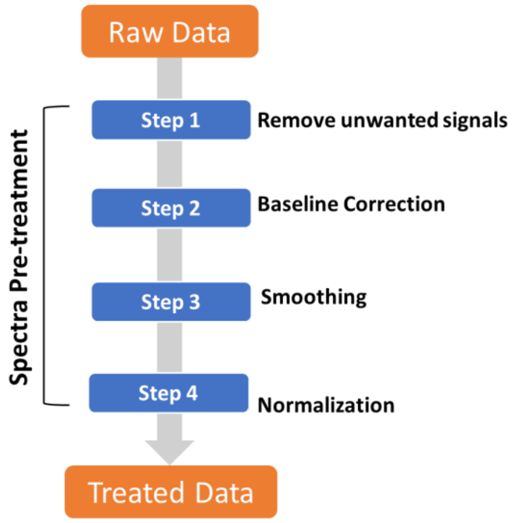
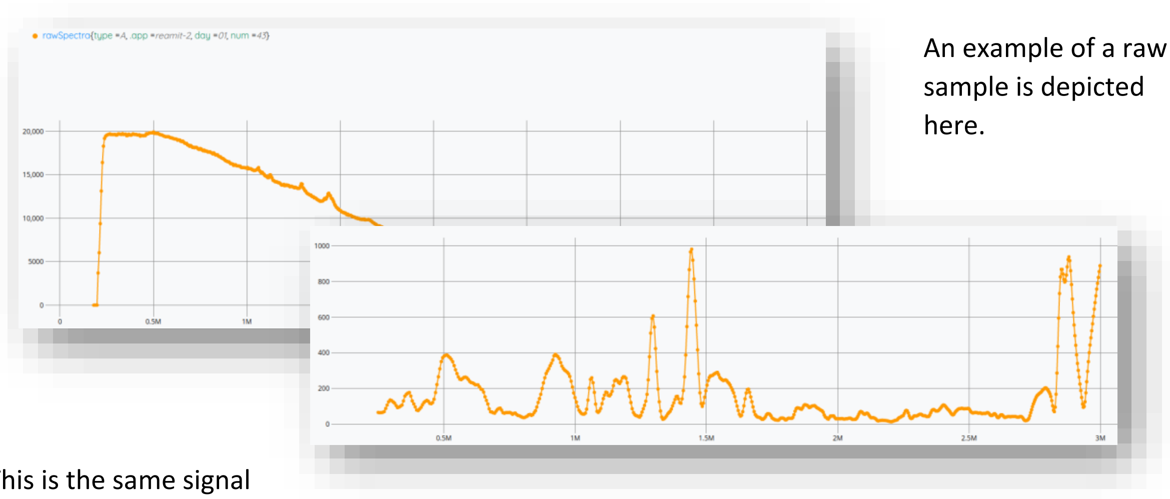


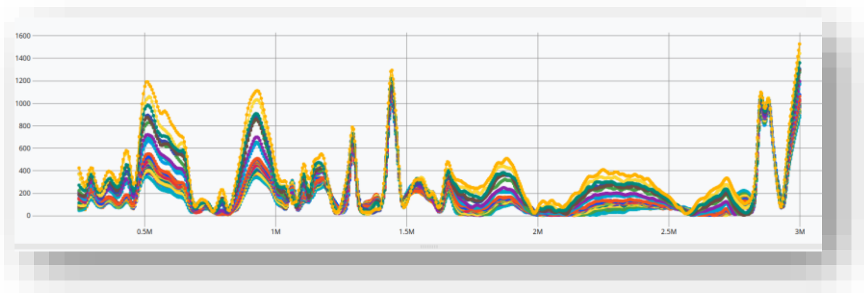
Fig. The necessary steps to pre-process Raman spectra

Analysis



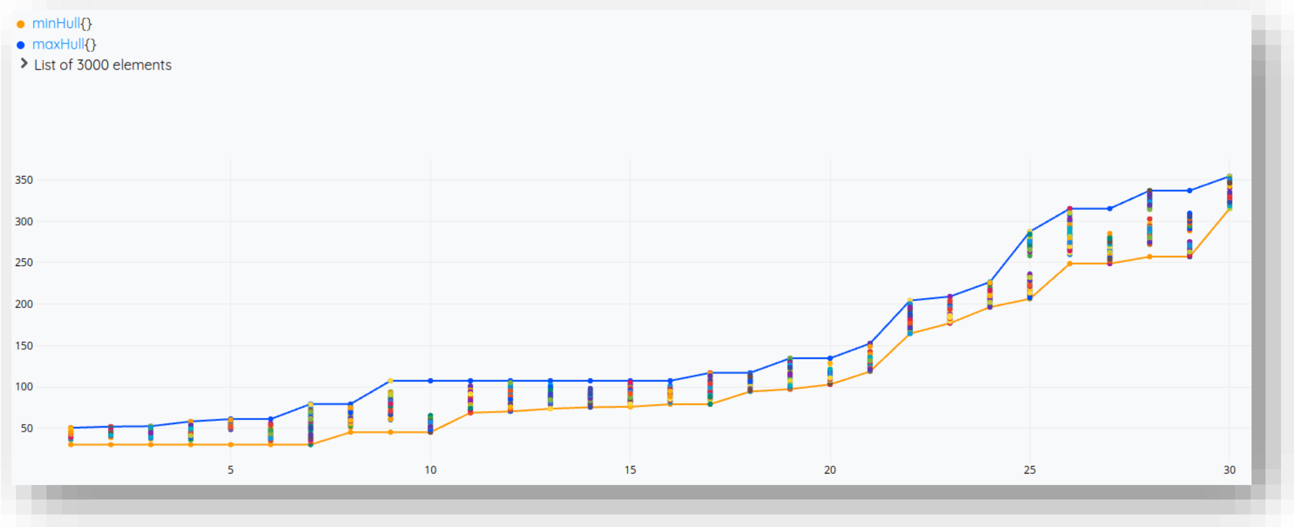
An example of a raw sample is depicted here.

This is the same signal after it is has been treated.

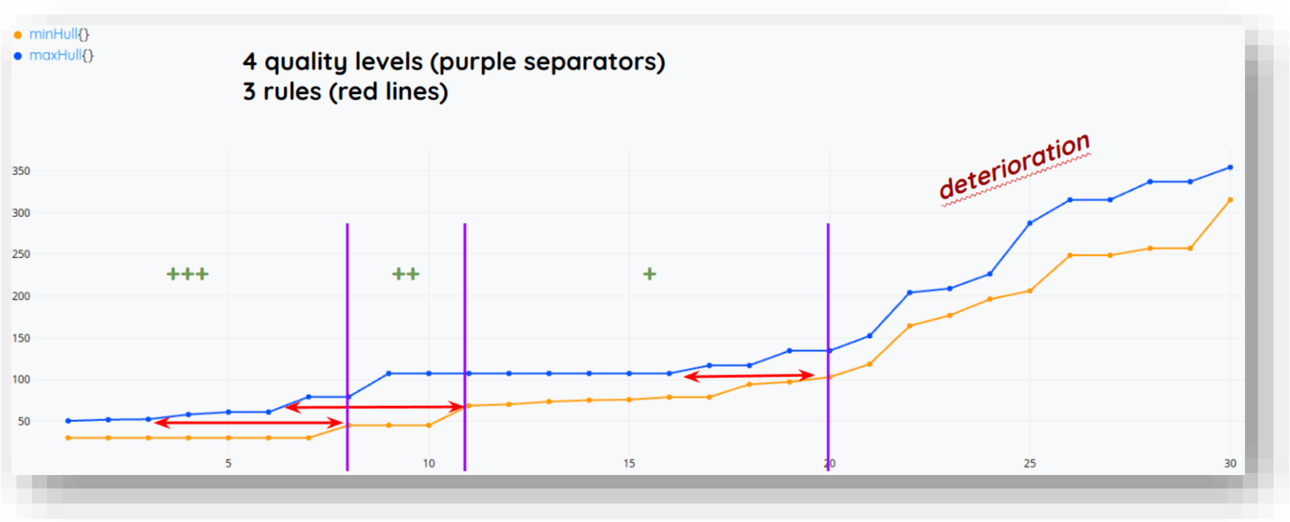


Here, we plotted the 30 day-mean series over the course of a month (each series aggregates 100 samples).

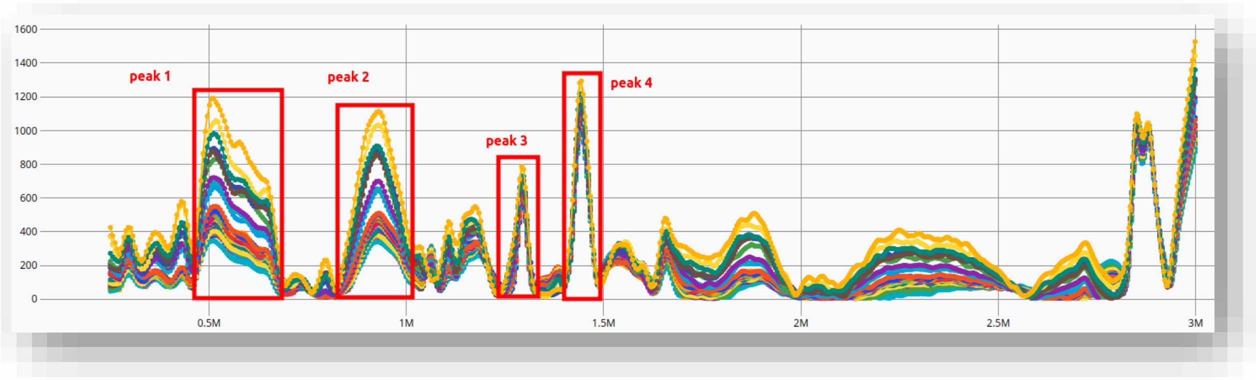
Next, we calculated the Euclidean distances of each sample from the mean series of the first day. In the following illustration, each point represents a result, with the day indicated on the x-axis. The blue and orange lines represent the upper and lower monotonic convex hulls, respectively.



As depicted in the figure below, the hull and the crossing points between its lower and upper lines can be utilized as boundaries to establish rules for an expert system. These rules can help define the quality level.



Even though the gist of the system given by those three rules, we have further enhanced the expert system by incorporating additional rules. By focusing on specific peaks and constraining the data, we have obtained a refined version of our model.



Interpretation

# samples	extra fresh	fresh	okay	expired
first model	384	513	774	1329
refined model	359	496	695	1450

Deployment

The 4-class chicken freshness model was successfully implemented and deployed on a local instance of the Warp10 platform running on a custom Raspberry Pi IoT module equipped with an LTE-M shield for remote connectivity during transportation. By utilising the Warp10 platform on the Raspberry Pi module, real-time analysis and classification of chicken freshness could be achieved. The classifier is designed to categorise chicken into four distinct freshness classes, "extra-fresh", "fresh", "good", and "expired", allowing for efficient monitoring and management of poultry inventory. The system architecture diagram in Figure x depicts the overview of the online chicken analysis system.

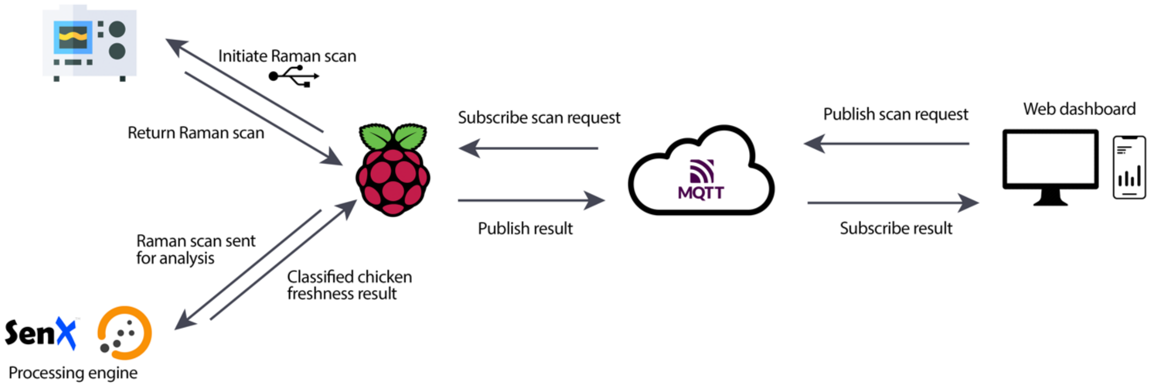


Figure: System architecture of the IoT Raman system for chicken classification.

To request a real-time scan and analysis, facilitate user interaction and provide a seamless experience, partners at Ulster University developed a frontend application. This application allows users to request Raman scans and, upon completion of the classification process, presents the results through a user-friendly interface. The system leverages MQTT (MQ Telemetry Transport), a lightweight, publish-subscribe, machine to machine network protocol for message queue/message queuing service. It is designed for connections with remote locations that have devices with resource constraints or limited network bandwidth, such as in the Internet of Things (IoT).

The frontend application was developed using PHP programming language and incorporates Bootstrap CSS, Chart.js, PHP-MQTT, and jQuery. To ensure security, the website requires a username and password for access, implemented through htaccess authentication. This authentication mechanism guarantees that only authorised users can initiate a scan and interact with the application, maintaining data integrity and confidentiality.

This interface enables end users to easily interpret and understand the freshness classification of the food samples. The user interface is accessible from https://finch.ulster.ac.uk/reamit

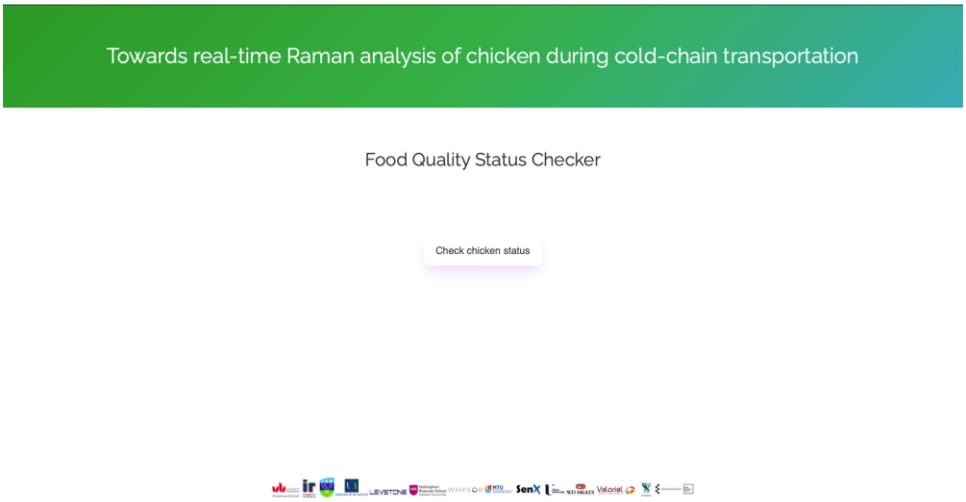


Figure: UI screen to initiate Raman scan.

Figure: Once the scan has been remotely triggered, scan data obtained, and classified using the Warp10 platform, the results are visualised on the front end using an easy to interpret freshness gauge for the end user. The status is also displayed as text below the visualisation.