Automated Fault Detection for a Distributed Fleet of Commercial Refrigerators

Haoyang (Kevin) Chen

Supervised by:

George Bell, Nick Ward

SKOPE Refrigeration

Submitted in part fulfilment of the requirements for the degree of
Master of Applied Data Science of the University of Canterbury
January 2020

Acknowledgements

I would first like to thank my colleagues at SKOPE Refrigeration: George Bell, Matthew Signal and Bill Mohs. They helped me to plan the project, answered my questions and provided feedback throughout this project. This project would not have been possible without them.

Many thanks to the previous intern of SKOPE Grace Cocks, for her excellent job. Her findings helped me build the fundamental parts of this project.

I would also like to thank my supervisor Nick Ward of the School of Mathematics and Statistics at the University of Canterbury, for his patience, enthusiasm and immense knowledge. He guided me from the basic concept and gave me invaluable advice during my project.

Contents

List of Figures

Chapter 1

Introduction

1.1 Project Background

SKOPE Refrigeration is a Christchurch based company that design and manufacture premium commercial refrigeration solutions. SKOPE's market-leading commercial refrigeration systems have recently joined the IoT movement. In 2018, SKOPE launched a product range fitted with the new Wellington Drive SCS Connect controller. A range of measures is monitored and collected by sensors fitted inside devices. These data are recorded at 30-minute intervals and then uploaded to a cloud server via a Bluetooth connected device. The recorded measures can be viewed and exported through SCS Connect Report Desktop app. The full list of available measures is in Appendix A. Customer can also monitor and customise fridge settings via SKOPE-connect mobile app.

When a fridge malfunction occurs, the customer calls SKOPE for repair service. Before the introduction of the controller solution, the technician relies on the malfunction description provided by the customer to prepare their service visit. Now the technician can use the past sensor data to help them diagnose the fault. An opportunity exists to use the collected data to enhance product development and improve service. It would be beneficial for the repair technician If we can detect fridge faults using the sensor data. The faults type we identified can help the technician to prepare the spare parts to bring when attending a service call. A greater understanding of the sensor data can give us more insight into the fridge working condition, such as usage patterns of different types of customers and how frequently each faults type occurs. These pieces of information can help SKOPE to improve their product design and adjust the testing parameters.

1.2 Project Objectives

There are two main objectives:

1. Define usage profiles of product and user groups to inform development testing.

   We want to know the usage profile of different user groups. Such as how the frequency of fridge door opens differs between restaurant and hotel, what ambient temperature does the refrigerator usually operate and in which way it affects the condenser temperature?

2. Utilise machine learning to develop fault detection algorithms based on sensor Data.

   Based on the sensor data and service call data, we can develop algorithms using R for automatic detection of faults. This automated detection would be helpful for the technician by ensuring they are prepared with the necessary parts when attending a service call. And also provide SKOPE with a greater understanding of the sensor data and faults, allowing SKOPE to identify whether sending a technician is necessary proactively.

Chapter 2

Methods

R package used: randomForest, tidyverse, gridExtra, viridis, shiny, ggplot2

2.1 Data Sources

The sensor and service record date of SKOPE refrigerators are managed by SKOPE and one of SKOPE's commercial partners.

1. Sensor data:

   The SCS software can export two types of files: a summary table of all connected fridges and sensor data of a single fridge. The summary table contains the serial number, outlet name, fridge model and other summary data such as total door swings and total operating hours.

   Sensor data are collected using sensors inside the refrigerators. This data is typically recorded as max, min and average values at 30-minute intervals and transmitted to a cloud server via the Bluetooth connected device.

   Sensor data can be viewed and exported from SCS Report desktop. There are 25 variables in the exported sensor data for each fridge, all in numeric format except the date. The number of fridges with sensor data available through the SCS Report desktop application is 923 for the SKOPE database and 2666 for the commercial partner's database.

2. Venue data:

   A table of customer ID, venue names and venue type. There are 535 customers with labelled venue type.

3. Service call data:

   The service records of the fridges managed by the commercial partner are store in multiple CSV files in the SKOPE internal website. The diagnosed fault types are recorded in a range of cause codes. The list of cause codes is available in Appendix B.

The service records of fridges managed by SKOPE stored in two separate database systems. We can only access the service record of a single fridge at a time by entering the serial number. The diagnosed faults are written in a paragraph of detailed descriptions.

2.2 Data Cleaning

Not all fridges with service record have sensor data. We first generate a CSV table of sensor data summary of all connected refrigerators via SCS software. We only select fridges with "In service" status.

The original service records from our commercial partner are stored in multiple CSV files. We first merge all CSV film into a single one, then we inner join this merged table with the serial number list of connected fridges. The result table contains the service record of connected fridges only. By applying a filter on the cause code column, we can find out the connected fridges with any particular faults types.This process is shown in Figure 2.1.

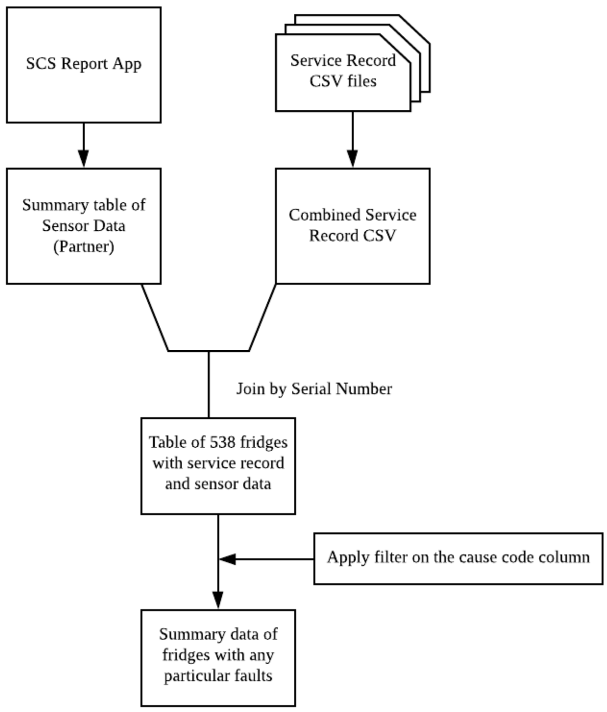

Figure 2.1: Flowchart of getting the list of connected fridges with service records

The raw sensor data are CSV files exported using the report app. The data is recorded in 30-min interval, but there are two rows for each timestamp. When reading the sensor data, we first combine the two rows of each timestamp into one. And then we convert the date format into “POSIXct” class in R.

Date	Ave powe	Door swin	Sensorles	Return air	Return air	Return air	Evap temp	Evap temp	Evap temp	Condenso	Condenso	Voltage m	Voltage m	Compress
10/04/2019 23:18	12	0	0	4.8	4	5				21	21	238	240	0
10/04/2019 23:18							0	0	1					
10/04/2019 23:48	508	0	0	3	2	5				34	36	234	240	100
10/04/2019 23:48							-5	-7	1					
11/04/2019 0:18							0	-7	0					
11/04/2019 0:18	120	0	0	3.5	1	4				31	36	238	246	5
11/04/2019 0:48							-4	-8	0					
11/04/2019 0:48	452	0	0	3.8	2	5				29	31	234	242	69
11/04/2019 1:18							-4	-8	0					
11/04/2019 1:18	356	0	0	2.7	1	4				29	31	234	240	49
11/04/2019 1:48							0	0	0					
11/04/2019 1:48	128	1	0	4.5	4	4				25	26	236	240	0
11/04/2019 2:18							-3	-8	0					
11/04/2019 2:18	392	0	0	3.6	2	5				29	32	234	240	56

Figure 2.2: An example of sensor data CSV file

2.3 Exploratory Data Analysis

2.3.1 Condenser Temperature

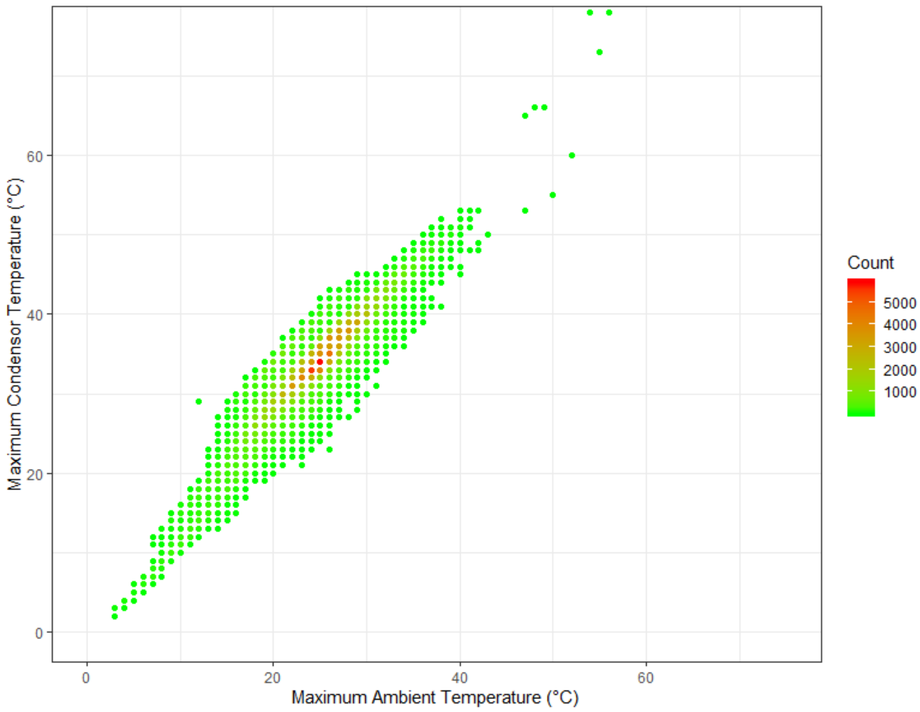

Figure 2.3: Scatter plot of the combination of maximum ambient temperature and maximum condenser temperature. The number of observations of each temperature combination are coloured in green-red colour gradient.

We sorted the fridge sensor data in SKOPE database by operation hours. Sensor data of 27 fridges with long operation hours were exported manually. There are 186823 observations in total.

Figure 2.3 is made using the combination of maximum ambient temperature and maximum condenser temperature from the 27 fridges. There is a positive relationship between the maximum of ambient and condenser temperature.

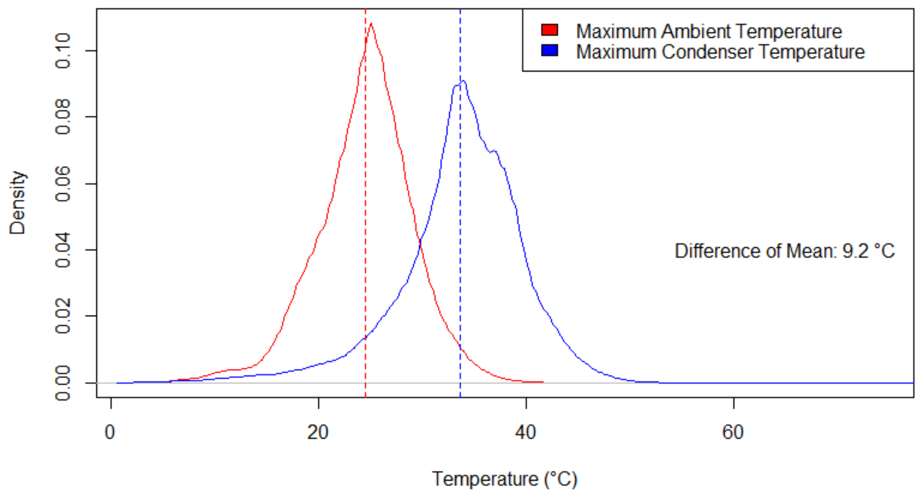

Figure 2.4: Distribution of maximum ambient temperature and maximum condenser temperature. The means of two distributions are marked using dashed lines.

Figure 2.4 shows that maximum ambient temperature is centred around 24 degrees Celsius, and the maximum condenser is centred around 33 degrees Celsius. The difference between the distribution means is 9.2 degrees Celsius.

2.3.2 Condenser Dirty

The condenser coils can be found at the back or at the top of the refrigerator depending on the model. Condenser allows heated refrigerants to get cooled. Dirty condenser retains heat, resulting in the compressor constantly stay on to maintain the set temperature.

Unlike condenser fan motor failure and evaporator freeze up, condenser dirty usually only cause performance issues and has no obvious visible outcomes for a short period of time. SKOPE is interested in how fast the condenser gets dirty and the amount of time before next maintenance is required.

Based on our previous findings in section 2.3.1, the condenser temperature is on average 9.2 degrees Celsius above the ambient temperature. Because of the seasonal change in ambient temperature. The condenser temperature decreases in the winter months and increases in the summer months. We would expect the difference between the ambient and condenser temperature increases overtime as dust accumulates on the condenser. We would also expect a step decrease in temperature difference and condenser temperature on the day of service visit when the condenser been cleaned. These patterns are shown in figure 2.5 below.

We can use the slope of the temperature difference as the rate of condenser get dirty. The data we need is long term sensor data of both ambient and condenser temperature probe. However, the fridge data in SKOPE database does not have long period of recorded sensor data. And the plotted temperature difference does not have a strong linear trend. The points are scattered around a horizontal line over a period of less than one year. My assumption is that maybe the measurement of ambient temperature is biased. The ambient temperature probe is installed inside the fridge. The emitted heat by the compressor and condenser may affect the readings. We choose to use the sensor data from the commercial partner's database for better data quality. However, most of the fridges in that database does not have an ambient probe. We can approximate the ambient temperature by using the minimum condenser temperature when the compressor is off for a full 30-minute window. When the compressor is off, the condenser temperature decay exponentially to the true ambient temperature. We calculate the mean of the condenser temperature when the compressor is off for each day to get the approximation of daily average ambient temperature.

However, minimum condenser temperature is also unavailable on the SCS report desktop application. A further approximation was made by using the average condenser temperature. This approximation is inaccurate as it is affected by the maximum temperature.

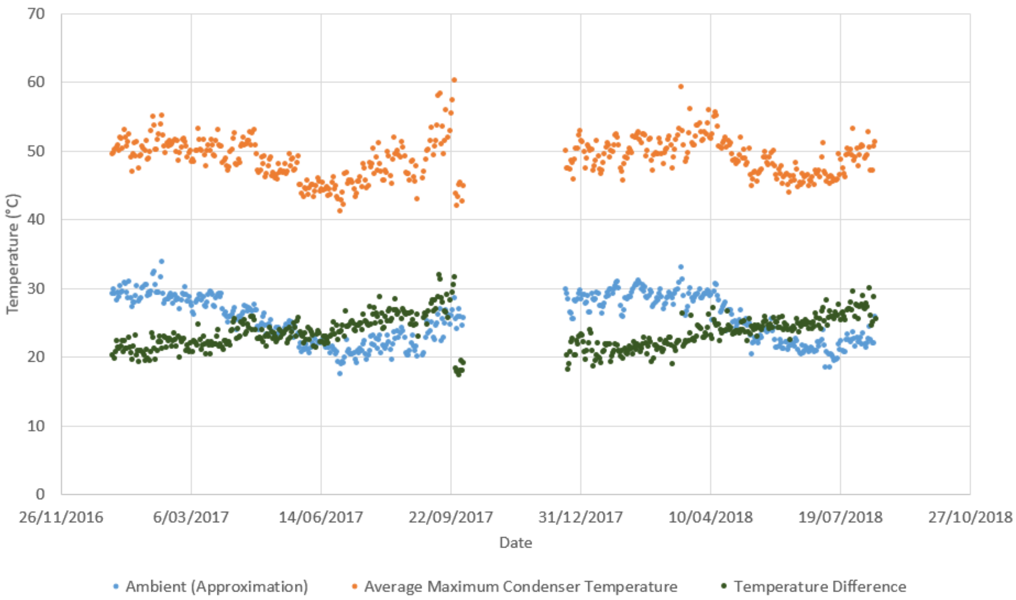

Figure 2.5: The Daily average of ambient temperature, difference between ambient (Approximation) and maximum condenser temperature for fridge with serial number 11X7987, this fridge was sent to repair on 24/09/2017 due to condenser dirty.

By using the approximated ambient temperature, we successfully identified a fridge with our predicted data patterns. As seen in figure 2.5, there is a clear seasonal trend in the ambient and condenser temperature, and a gradual increase in the temperature difference as more dust and debris settles on the coils. A service record belongs to this device was created on 24/09/2017 with faults type been condenser dirty. This fridge had been out of service for 3 months for repair, and we see a steep decrease in the temperature difference after the repair.

Figure 2.6 shows the temperature difference between ambient and condenser of three fridges. The temperature difference increases over time, and the relationship is roughly linear. The increasing rate is about 0.02 degrees Celsius per day or 2 degrees per 100

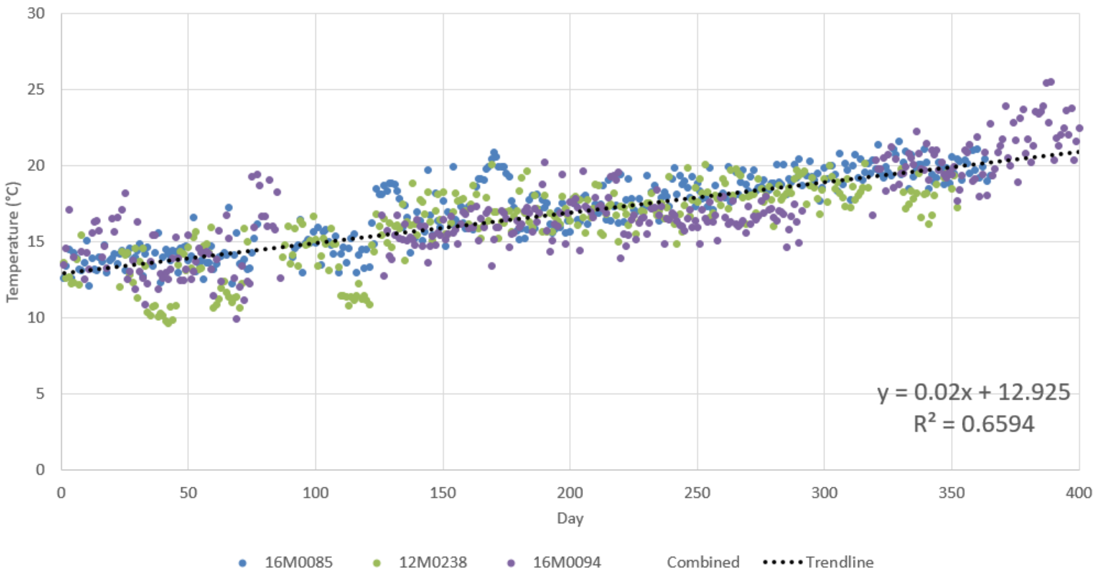

Figure 2.6: Daily average difference between ambient (Approximation) and maximum condenser temperature of three fridges .

days. But these three fridges are just a small subset of the whole dataset. Most of the fridge data are either too fragmented or have too short time period. For those fridges with continuous data records, most of them show no sign of significant trends. For those fridges with an increasing trend in temperature difference, the increasing rate range from 1 degree per 100 days to 3.5 degrees per 100 days as shown in figure 2.7.

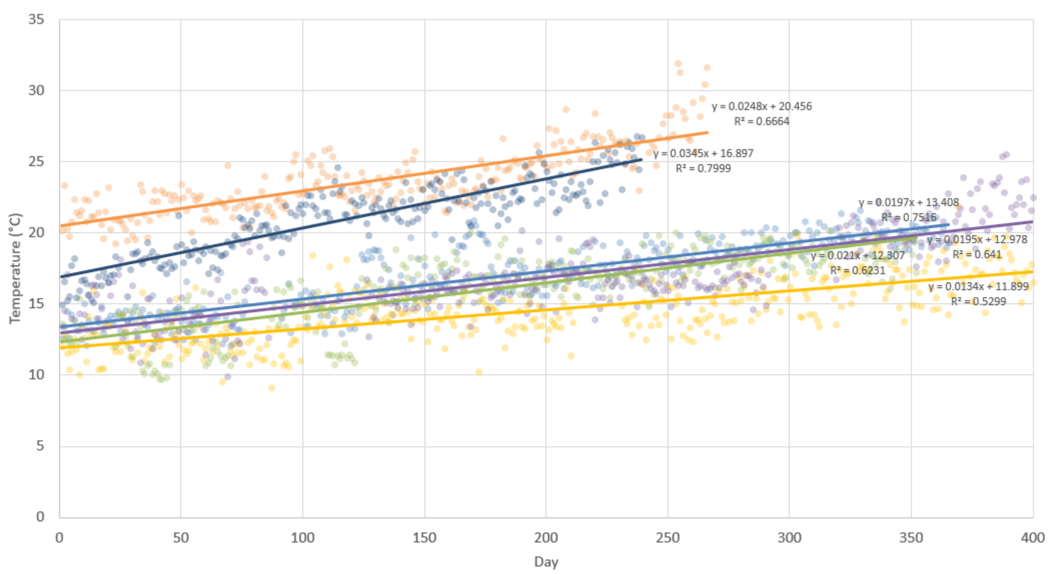

Figure 2.7: Daily average difference between ambient (Approximate) and maximum condenser temperature of selected fridges

2.3.3 Door Swings

We exported the sensor data of 60 fridges that have at least 50 days of records. These fridges are from 10 venue categories. We calculated the daily average of maximum ambient temperature and the total number of door swings each day. The data of these 60 fridges are in different length. To avoid the bias effect from any single fridges, we select 30 random rows from each fridge.

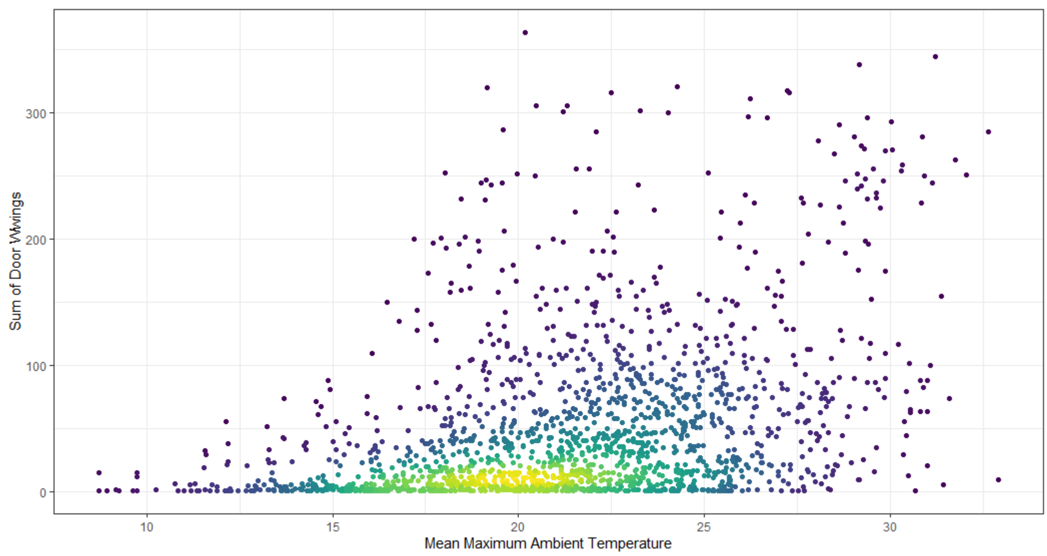

Figure 2.8: Scatter plot of average maximum ambient temperature vs daily door swings

As shown in figure 2.8, we plotted the combination of the daily average of maximum ambient temperature and the total number of door swings each day in a scatter plot. The density is represented using Viridis colour scales (Yellow to purple). There is no significant linear relationship between these two variables. We will have heteroscedasticity problem if we try to train a linear model on it.

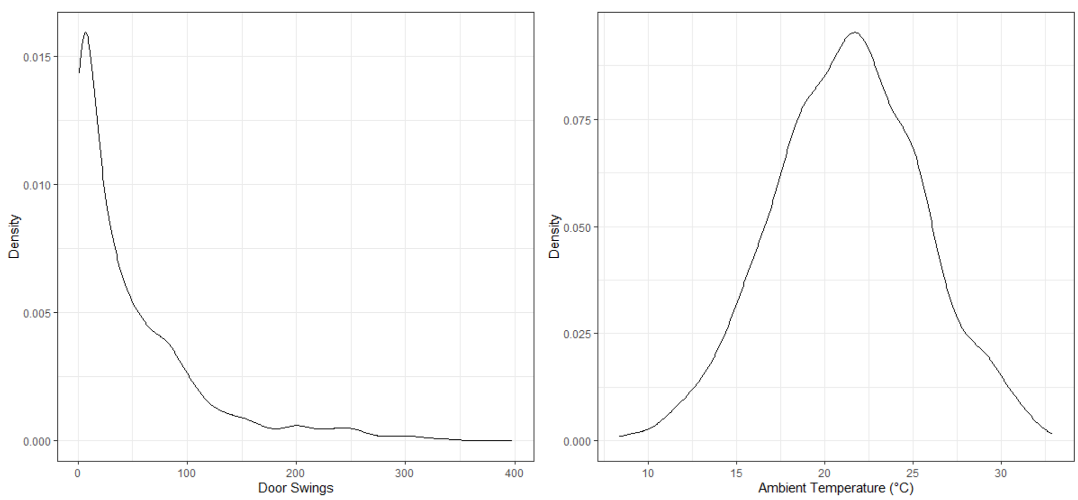

Figure 2.9: Left: Distribution of daily door swings; Right: Distribution of average ambient temperature

Figure 2.8 suggested that most fridges have less than 50 door swings each day, and the most common ambient temperature is around 20 degrees Celsius. This finding is supported by figure 2.9.

Because door swings is an indicator of fridge usage, we made figure 2.10 and 2.11 to study the difference in usage patterns among different types of customers.

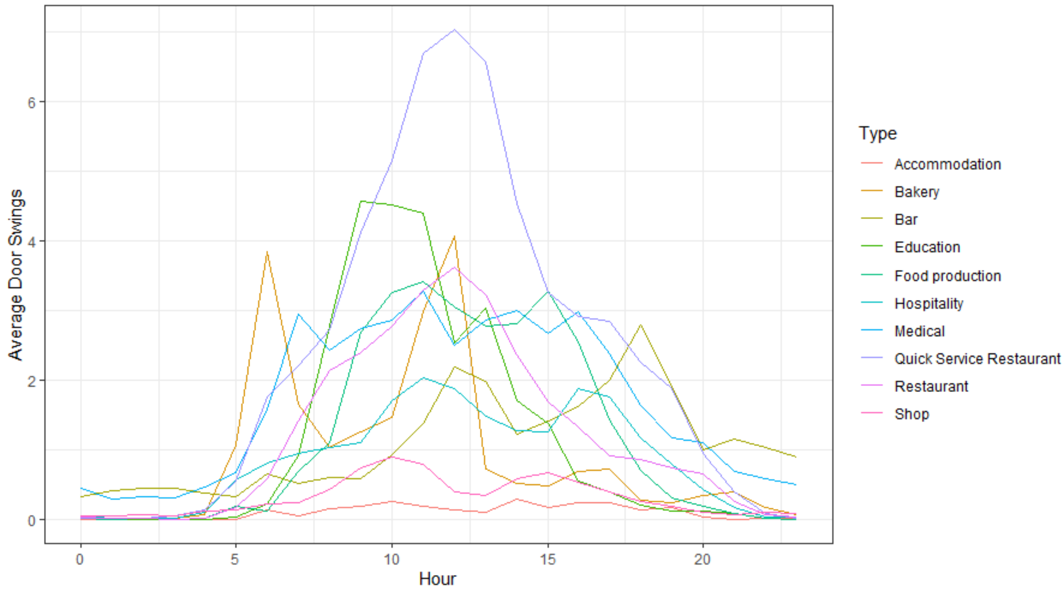

Figure 2.10: Line chart of average door swings during the day by venue type.

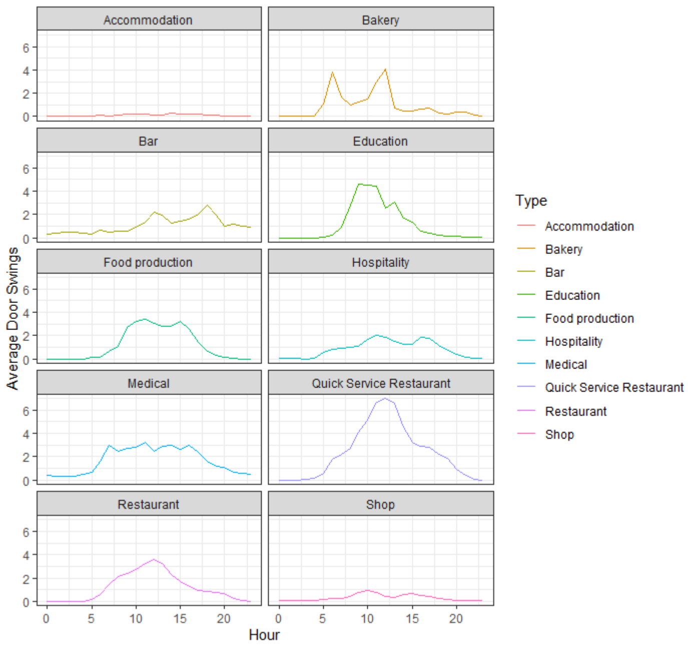

Figure 2.11: Facet line chart of average door swings during the day by venue type.

Figure 2.10 shows that customers of all venues types tend to use the fridges more often during the midday. Quick service restaurant has the highest number of door swings. Figure

2.11 shows that fridges in bars usually been opened during the evening hours. Fridges in bakeries have two time periods of high door swings: around 7 am and midday.

One of the most common fault types in the service records is Door fault. We identified 31 service records that have "door" as their fault code.

There are 1385 services records after cleaning. 35 of them are door faults.Therefore the overall percentage of door fault is 2.5 %. We separate the fridges into six brackets based on the number of total door swings. The percentage of door faults for each bracket is calculated by divide the number of door fault fridges by the total number of connected fridges. Then we divide this percentage by the overall percentage to get the risk. For example, there are 688 connected fridges with less than 500 door swings and 1 of them has a door fault record. The percentage for fridges with less than 500 door swings is 1/668, 0.15%. The risk of door fault for fridges with less than 500 door swings is 0.15%/2.5% = 6%. And we can interpret this result as fridges with less than 500 total door swings are 94% less likely to have door faults. We also studied the relationship between door faults and door swings per day. Door swings per day are calculated by dividing the total door swings by the total operating hours. We found that the risk of door fault increased rapidly for fridges with more than 25 door swings per day or 5000 door swings total.

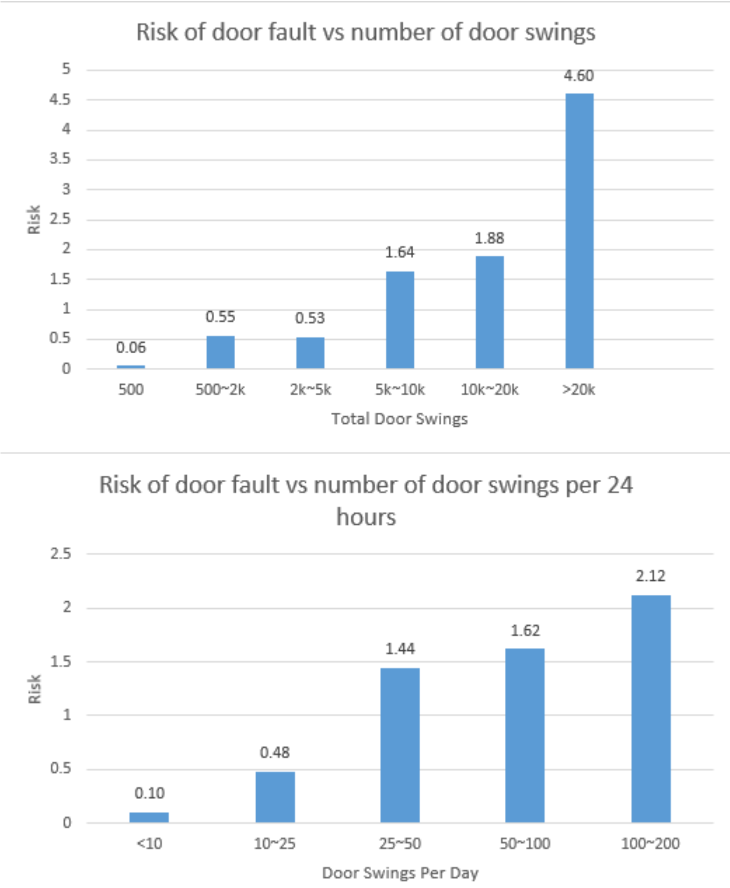

Figure 2.12: Top: Risk of door fault vs number of door swings; Bottom: Risk of door fault vs number of door swings per 24 hours.

2.3.4 Compressor On Time

Another way to study the usage pattern of fridges in difference venues is by plot the distribution of compressor on time for each venue type.

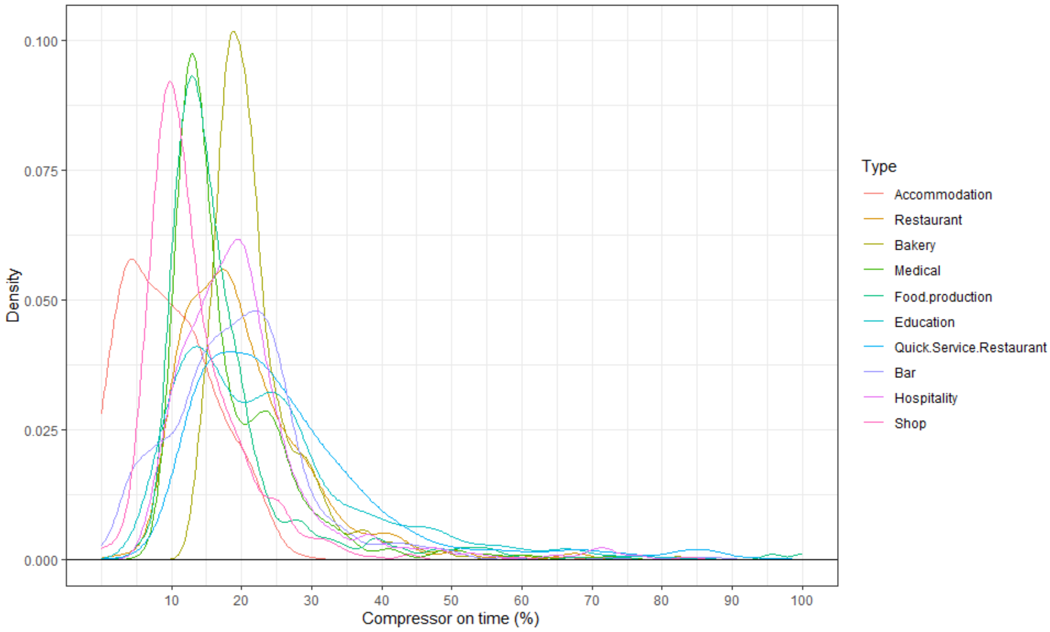

Figure 2.13: Distribution of compressor on time for difference venue types

The shapes of distributions for different venue types are similar. Fridges from all venue types have a compressor on time of around 20% for most of the time.

2.4 Statistical Learning Method

We want to know which predictors are impotent in fault detection. The machine learning we choose is random forest.

Random forest is a tree-based statistical learning method. This method partition the predictor space into a number of regions, the set of splitting rules can be summarized in a tree consisting of nodes and branches. All predictors are considered using a cost function at each split points. The split with the lowest cost is selected [James et al., 2017]. We want to find the partition rules that can minimize the RSS of the tree, but it is computationally infeasible. Instead, we use an approach called recursive binary splitting. We recursively split the predictor space where each split defines two new branches further down the tree. This approach is greedy because we only look for a local optimal at each split.

Let's say we split the predictor space into two regions R_1 and R_2. The predictor we selected is X_j, and the cutpoint is s.

$$R_1(j, s) = \{X|X_j < s\} \;\; and \;\; R_2(j, s) = \{X|X_j >= s\} \tag{2.1}$$

and we seek the value of j and s that minimize the equation

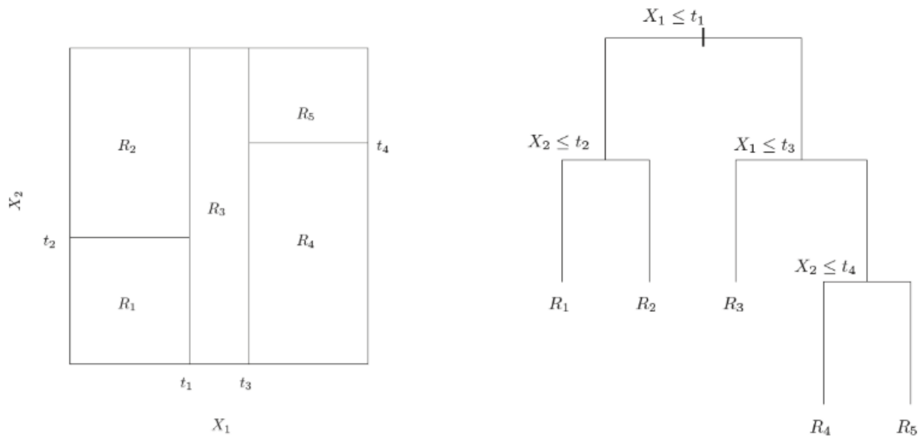

Figure 2.14: Left: A recursive binary splitting on a two-dimensional example. Right: A tree corresponding to the partition of the left example [James et al., 2017].

$$\sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{2.2}$$

To predict the response for a given test observation, we use the mean of the response values of the training observations in the region to which that test observation belongs.

The traditional decision tree has high variance issues and can be very non-robust. This means that a small change in the data can cause a large difference in the results. One procedure to lower the variance of our learning method is called bootstrap aggregation or bagging. Because we usually do not have access to multiple training sets, we resample the training data and construct trees using the resampled training sets. Each individual tree is deep and low bias. We average all prediction results from those trees to reduce the variance [James et al., 2017].

The predictions from the bagged trees are usually very similar due to some strong predictors. Bagged trees will always use these strong predictors first. Random forest decorrelating the trees by forcing each split to choose one predictor from a random sample of m predictors. The number of predictors (m) considered at each split is usually set as the square root of the total number of predictors (p) for classification and p/3 for regression. Only a subset of predictors is considered so other weaker predictors can have more chance [James et al., 2017].

2.5 Fault Detection

2.5.1 Evaporator freeze up

When evaporator freeze up occurs. We would expect a significant decrease in the evaporator temperature. The sensor data of three fridges with service record of evaporator freeze up was exported using the SCS Connect Report app. The exported data is in CSV form and start from the week before the create data of their service record. A new column called Freeze was added manually to each CSV file as the response variable. The Freeze response range from 0 to 1. With 1 means confirmed free up occurred and 0 means no freeze up. This manual labelling process is based on the minimum temperature of the evaporator.

This two csv is joined to form a complete training dataset. By checking the importance of these variables (Figure 2.16), we found that the most important one besides evaporator temperature is the return air temperature and compressor on time. We plotted the return air temperature and the average evaporator temperature of a single fridge using SCS software. We can see that the Compressor on Time increases to 100% frequently, only decreasing for defrost cycles.And the Return Air Temperature increases before settling at a maximum value, even with the compressor on 100%.

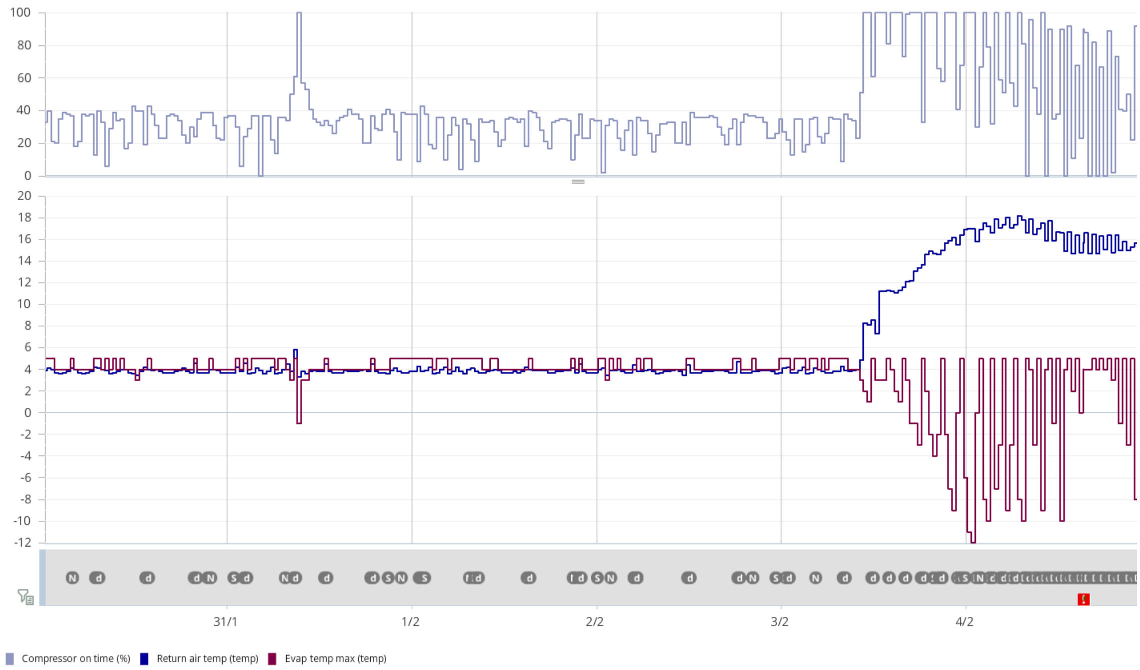

Figure 2.15: SCS graphic output of the freeze up period

We also noticed that the data tend to fluctuate a lot. To minimise the effect of this fluctuation, we need to implement a smoothing method on the data. We also want to take into account the sequential nature of these time series data to let the past data can have an effect on the future.

The method we took is called Single Exponential Smoothing.Exponential Smoothing explicitly uses an exponentially decreasing weight for past observations. Specifically, past observations are weighted with a geometrically decreasing ratio.

Single Exponential Smoothing requires a single hyperparameter, alpha. Alpha is set to a value between 0 and 1. A value close to 1 indicates that the model is fast learning and pays more attention to the recent observations, whereas smaller value close to 0 means the model let the past observations to have a larger effect [Brownlee, 2019].

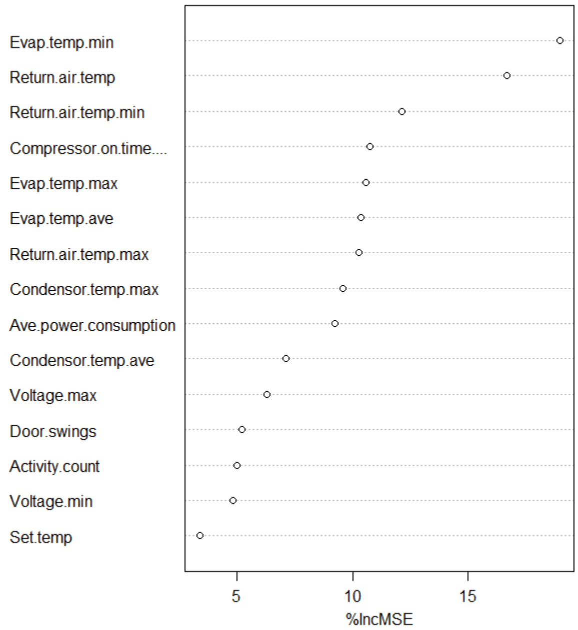

Figure 2.16: Importance ranking for evaporator freeze up

The basic formula is:

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1} \tag{2.3}$$

Where S_t is the smoothed value at time t [Hyndman and Athanasopoulos, 2013].

The best value of alpha hyperparameter may be different for each predictor. The alpha value must range from 0 to 1. I set three alpha values (0.25, 0.5 and 0.75) for three important predictors: minimum evaporator temperature, compressor on time and return air temperature.

We found that the best alpha for minimum evaporator temperature , compressor on time and return air temperature is 0.25,0.75 and 0.75 respectively (Figure 2.17). We then remove the columns of unselected alphas and train a random forest model on it.

Compared to the model without exponential smoothing variables (Figure 2.18), the MSE of our final model decreases from 0.005 to 0.0043, and the percentage of variance explained increases from 97.37% to 97.73. This suggests that the exponential smoothing slightly improved our model quality.

An issue with our data is that there is a large proportion of fridges does not have an evaporator temperature probe. Therefore, a second model was trained based on data without evaporator temperatures using the same steps above.

The code for freeze up detection and its dependent function has been wrapped up into a single function. The freeze up detection function will first choose the correct model based on the availability of evaporator temperature data. The function will find the time when the freeze-up response goes higher than 0.9. It also checks the data from the last 2 days

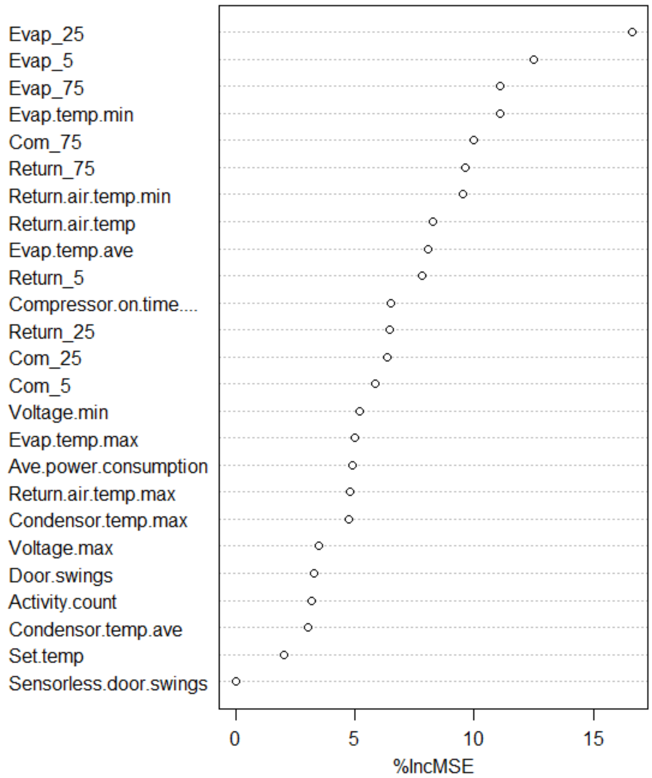

Figure 2.17: Importance ranking for evaporator freeze up with additional smoothing variables

```
> print(rf.pre)

Call:
 randomForest(formula = Freeze ~ ., data = Train_df[, -1], ntree = 200,      importance = TRUE)
               Type of random forest: regression
                     Number of trees: 200
No. of variables tried at each split: 5

          Mean of squared residuals: 0.005031974
                    % Var explained: 97.37
> print(rf.train2)

Call:
 randomForest(formula = Freeze ~ ., data = rf_train_EMA2[, -1],      ntree = 200, importance = TRUE)
               Type of random forest: regression
                     Number of trees: 200
No. of variables tried at each split: 6

          Mean of squared residuals: 0.004332933
                    % Var explained: 97.73
```

Figure 2.18: Up: without exponential smoothing variables, Down: with exponential smoothing variables

at a lower sensitivity in order to find ongoing freeze-ups.

The function will generate three types of text outputs:

- Evaporator Freeze up not detected: No sign of freeze up based on historical data.
- Evaporator freeze up detected at . . . : The time when freeze up was detected.

- High Chance of evaporator freeze up: High change of ongoing freeze up based on the last 2 days of data (Lower sensitivity).

2.5.2 Condenser Fan failure

When a condenser fan fault occurs, less air being blown over the coils and the condenser temperature will increase to its maximum value in a very short time period as hear being removed at a lower rate.

We identified 18 service records with condenser fan motor fault code. But only 5 of them is suitable as our training or testing data. The rest of them are has no sensor reading records during the service visit period.

The sensor data of fridges with a service record of condenser fan failure was exported using the SCS Connect Report app. The exported data is in CSV form and start from the week before the create data of their service record. We select one fridge data file and add a new column called Fan as the response variable. The Fan is a binary response variable of 0 or 1. With 1 means confirmed fan failure occurred and 0 means normal. This manual labelling process is based on the maximum temperature of the condenser temperature.

The importance test shows that the most important predictors (Figure 2.20) are condenser temperature maximum, Return air temperature and evaporator temperature maximum. By plotting these three variables using SCS software (Figure 2.19), we found that the maximum condenser temperature experiences a sudden step increase to its maximum value. And both the return air temp and evaporator temp begin gradually increasing when the condenser temp spikes up. The method is similar to what we used for evaporator freeze up is applied. But because the normal condenser temperature of other fridge models may differ from the normal condenser temperature of the fridge we used as training data. Therefore, we adjust the model by comparing the smoothed values with its means. The prediction output is a numeric response variable range from 0 to 1. To further reduce the false positive rate, a fridge is classified as condenser fan failure when the predicted response is larger than 0.9 for 4 continuous timestamps. The statistical summary (Figure 2.21) shows that our model explained 99.25% of the variance.

2.6 Shiny App

To launch the Shiny app, first, open the detect.R file with RStudio. Then set the working directory to the source file location and enter the following line of code.

```
runApp("detect.R")
```

The interface (Figure 2.22) of this app consists of three parts: File upload box on the left-hand side, text and graphic output on the right-hand side. The input file has to be the CSV format sensor data as exported from the SCS report app. The text output will show whether condenser or evaporator faults have been detected and the timestamp of the faults. The graphic output area will display line charts of the important predictors versus time for the detected fault types. When no fault been detected, the graphic area will be empty and display two line charts when both faults been detected

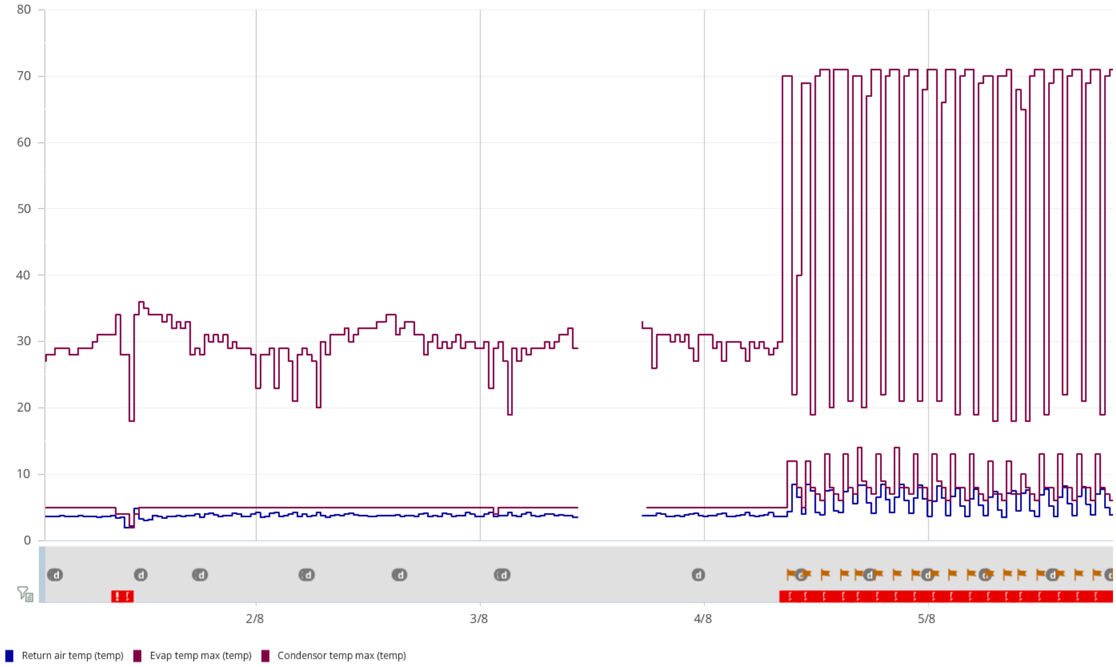

Figure 2.19: SCS graphic output of a fridge with condenser fan failure

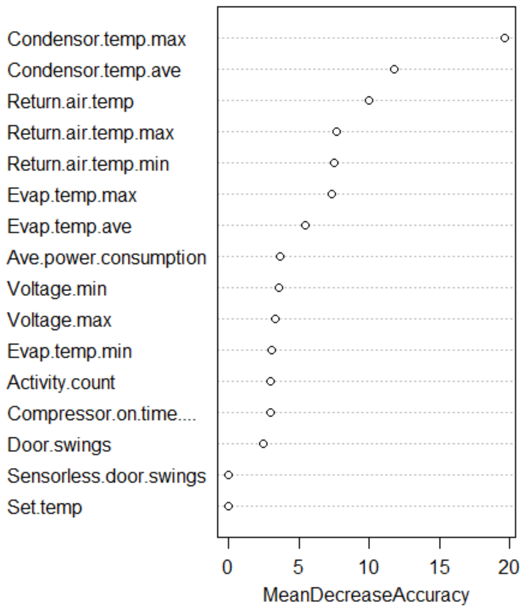

Figure 2.20: Importance ranking for condenser fan failure

```
Call:
 randomForest(formula = Fan ~ ., data = rf_fan_DIFF[, c(18, 22:24)],      ntree = 200, importance = TRUE)
               Type of random forest: regression
                     Number of trees: 200
No. of variables tried at each split: 1

          Mean of squared residuals: 0.001307205
                    % var explained: 99.25
```

Figure 2.21: Statistical summary of condenser fan failure model

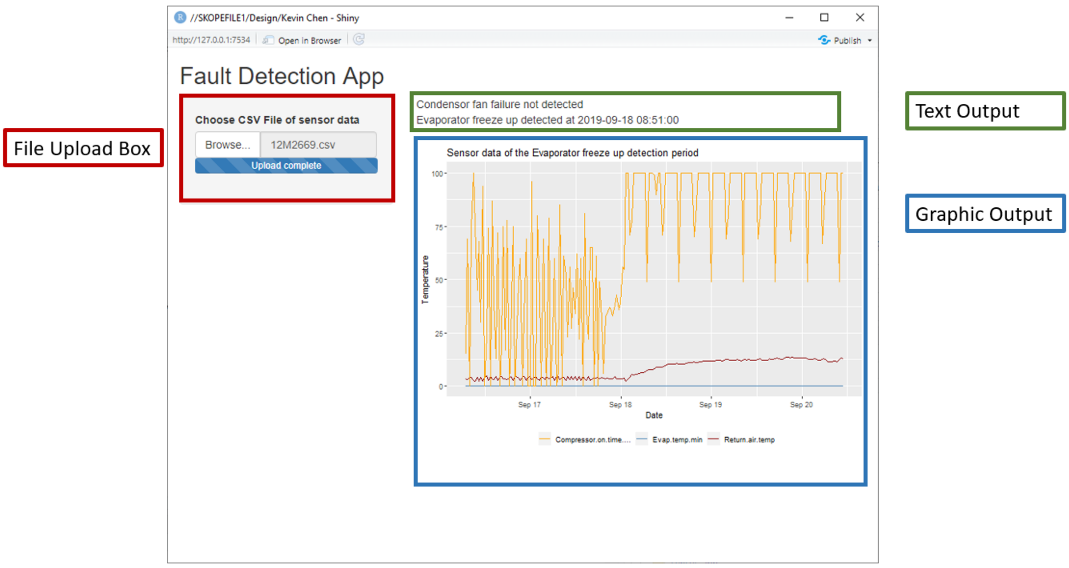

Figure 2.22: The Interface of the Fault Detection App

Chapter 3

Results

3.1 Key Findings of Exploratory Data Analysis

1. The average maximum ambient condenser temperature is 24.5 degrees Celsius.

2. The average maximum condenser temperature is 33.7 degrees Celsius.

3. There is a positive relationship, and the difference of mean is 9.2 degrees Celsius.

4. The temperature difference between ambient and condenser increases over time, and the relationship is roughly linear.The increasing rate range from 1 degree per 100 days to 3.5 degrees per 100 days.

5. There is no significant linear relationship between ambient temperature and number of door swings.

6. Most fridges have less than 50 door swings each day, and the most common ambient temperature is around 20 degrees Celsius.

7. Customers of all venues types tend to use the fridges more often during the midday.

8. Quick service restaurant has the highest number of door swings. Fridges in bakeries have two time periods of high door swings: around 7 am and midday

9. The risk of door fault increased rapidly for fridges with more than 25 door swings per day or 5000 door swings total.

10. Fridges from all venue types have a compressor on time of around 20% for most of the time.

3.2 Model Validation

We use data of both the SKOPE and commercial partner for validation. Because the service records always created after the actual faults occurred. Therefore, we expected the fault date detected by the model a few days before the service record creation date.

3.2.1 Evaporator freeze up

We exported sensor data of 7 fridges with evaporator freeze up faults from our commercial partner's database and 6 fridges with no evaporator freeze up service record as our testing data to validate our model. The testing results show that all fridges with faults have been detected. And there is 1 false positive for the normal fridge data, the false positive rate is 1/6.

Serial Number	Service Record Creation Date
15M1017	23/01/2019
15M1186	3/12/2018
16M0129	17/12/2018
16M0432	16/10/2018
16M0447	4/02/2018
18M1779	2/01/2019
18M1782	3/04/2019

Table 3.1: Serial Number and service record creation date of testing data (Evaporator)

```
L:/Kevin Chen/Evaporator_Freeze_Up/Test/15M1017.csv    Evaporator freeze up detected at 2019-01-21 09:49:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/15M1186.csv    Evaporator freeze up detected at 2018-12-03 11:46:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/16M0129.csv    Evaporator freeze up detected at 2018-12-16 22:02:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/16M0432.csv    Evaporator freeze up detected at 2018-10-16 19:23:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/16M0447.csv High Chance of evaporator freeze up 2018-02-03 23:54:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/18M1779.csv    Evaporator freeze up detected at 2018-12-31 13:12:00
L:/Kevin Chen/Evaporator_Freeze_Up/Test/18M1782.csv    Evaporator freeze up detected at 2019-03-17 12:08:00
L:/Kevin Chen/Condenser_Fan/Test/skope/HA180917603.csv                    Evaporator Freeze up not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181018549.csv                    Evaporator Freeze up not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181219903.csv                    Evaporator Freeze up not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181219986.csv                    Evaporator Freeze up not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA190210439.csv                    Evaporator Freeze up not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA190714501.csv Evaporator freeze up detected at 2019-10-30 10:20:00
```

Figure 3.1: Validation outputs (Evaporator)

3.2.2 Condenser fan fault

We use the remaining datasets of 4 fridges with condenser fan fault to validate our model.

Serial Number	Service Record Creation Date
12M1138	1/11/2018
13M3856	15/04/2019
13M3973	7/04/2019
13M4123	24/07/2018

Table 3.2: Serial Number and service record creation date of testing data (Condenser fan)

Our model successfully detected the condenser fan failure and the correct date for all four testing data.

```
> result

L:/Kevin Chen/Condenser_Fan/Test/12M1138.csv Condensor fan failure occurs at 2018-11-01 08:55:00
L:/Kevin Chen/Condenser_Fan/Test/13M3856.csv Condensor fan failure occurs at 2019-04-13 15:50:00
L:/Kevin Chen/Condenser_Fan/Test/13M3973.csv Condensor fan failure occurs at 2019-04-07 11:56:00
L:/Kevin Chen/Condenser_Fan/Test/13M4123.csv Condensor fan failure occurs at 2018-07-23 13:26:00
```

Figure 3.2: Validation outputs 1 (Condenser fan)

We also exported 6 fridge data from the SKOPE database to further testing the model. 3 out of 6 fridges has been detected as had condenser fan failure. By manually checking the SKOPE service record database (Figure 3.4), we found that only 1 of them is correct while the other 2 are false positive results. No repairs history has been found for the 3 negative results.

```
L:/Kevin Chen/Condenser_Fan/Test/skope/HA180917603.csv Condensor fan failure occurs at 2018-12-21 15:10:00
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181018549.csv                 Condensor fan failure not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181219903.csv Condensor fan failure occurs at 2019-08-21 14:35:00
L:/Kevin Chen/Condenser_Fan/Test/skope/HA181219986.csv Condensor fan failure occurs at 2019-10-02 16:33:00
L:/Kevin Chen/Condenser_Fan/Test/skope/HA190210439.csv                 Condensor fan failure not detected
L:/Kevin Chen/Condenser_Fan/Test/skope/HA190714501.csv                 Condensor fan failure not detected
```

Figure 3.3: Validation outputs 2 (Condenser fan)

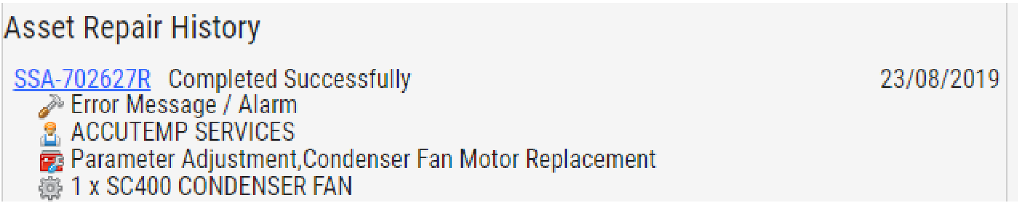

Figure 3.4: Service record of HA181219903 viewed via SKOPE service record system

Chapter 4

Discussion

Because of the difficulties in data export, the size of our testing data is very limited. And since the SCS report does not allow multiple sensor data to be exported at once, all sensor data have to be manually exported from the SCS report desktop application one by one. The time period of exported data can only be determined manually for each fridge. This process is very time-consuming.

Another issue in our model is that we did not include the ambient temperature, because a large proportion of fridges in the commercial partner's database does not have ambient probe installed.

The validation results show that our models have high sensitivity when detecting fridges with condenser fan failure or evaporator freeze up. But the false positive rate is high. Our models can only detect one type of fault. Because the sharing pattern between faults, fridges with another fault can cause incorrect detect results. For example, the condenser fan failure and evaporator freeze-up share the common characteristic of the increase in return air temperature.

The sample size of our training and testing data is also very limited due to the small population of avaliable data. There are a large amount of invalid data, and most faults types only have less than 5 fridges of data after applying necessary filters.

Chapter 5

Conclusion

Upon completion of this project, we know that using our random forest model to detect fridge faults is possible but not reliable. To improve our model, we require more data to train and test the model. Condenser fan failure and evaporator freeze up can be detected by our model but with a high false positive rate.

Bibliography

[Brownlee, 2019] Brownlee, J. (2019). A gentle introduction to exponential smoothing for time series forecasting in python. https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python.

[Hyndman and Athanasopoulos, 2013] Hyndman, R. J. and Athanasopoulos, G. (2013). Forecasting: Principles and practice. https://otexts.com/fpp2/ses.html.

[James et al., 2017] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). *An introduction to statistical learning*, volume 112. Springer.

Chapter 6

Appendix

6.1 Appendix A: List of Measures available on SCS Report Desktop App

- Activity Count
- Ambient Temperature
- Average Power Consumption
- Compressor on Time (%)
- Compressor Power
- Condenser Temperature (Max and Average)
- Door Swings
- Evaporator Power
- Evaporator Temperature (Min, Max and Average)
- Lighting 1 Power
- Return Air Temperature (Min, Max and Average)
- Sensorless Door Swings
- Set Temp
- Voltage (Min and Max)

6.2 Appendix B: List of Cause Codes

- Compressor Capacitor/Relay Fault
- Compressor Fault
- Condenser Dirty
- Condenser Fan Motor
- Control Board Fault
- Deck Failure
- Door
- Door Glass
- Door Sensor
- Drain Blocked / Leak
- Electronic Control Fault
- Electronic Probe (Non EMS Control)
- EMS Controller
- EMS Door Switch
- EMS Probe
- Environment Outside Specifications
- Equipment Missing
- Evaporator Fan Motor
- Evaporator Freeze Up
- Faulty Power Lead
- Handle
- Light Failure
- New Equipment Failure

- Night Curtain Problem
- No Access to GPO
- No Access to Unit
- No Fault Found
- No Power
- Operator Fault
- Overloaded Power Circuit
- Panel
- Refrigeration Leak
- Refurb Equipment Failure
- Sensors
- Software Issue
- Switch Door
- Switch Faulty
- Switch Power
- Temp Control
- Thermal Protection Cut out
- Transformer
- Wiring Issue