Failure Detection for Pharmaceutical Cold Chain Logistics

Envirotainer

By

Yun Huang

Jingyun Jiang

Duo Zhou

Supervisor: Ashish Pujari

A Capstone Project

Submitted to the University of Chicago in partial fulfillment
of the requirements for the degree of

Master of Science in Analytics

Graham School of Continuing Liberal and Professional Studies

05, 2021

Abstract

As a cold chain solution company that supports pharmaceuticals transportations, Envirotainer's objective is to successfully maintain cargo containers' set temperature while delivering temperature-sensitive pharmaceuticals. This project uses descriptive analysis to generate a failure risk profile based on Envirotainer's historical container logged data and shipment data to understand the characteristics of failure causes. Temperature deviations, performance anomalies, and container parts malfunctions are defined as shipment failures. In addition, different machine learning methods are utilized to develop a best model to predict the likelihood of failure prior to a shipment given certain information. With the recommendations and insights generated from both descriptive and predictive analyses, Envirotainer can reduce shipment failure by taking preventative measures on top failure causing factors and improving container handling training.

Keywords: pharmaceuticals, cold chain, logistics, failure risk profile, anomaly and failure prediction

Executive Summary

Since each shipment contains high-value pharmaceuticals, Envirotainer's goal is to minimize shipment failure rate while maintaining desired container condition. A failure risk profile is developed in order to reach that goal. Failures are identified as temperature deviations from a set temperature range, essential parts failure alarms, and container compressor performance anomalies. Certain shipment conditions and container handling behaviors are identified as the top shipment failure causes. Shipments are more likely to experience temperature deviations and compressor/heater failures during summer. Likewise, the temperature deviation rate increases when Delta (average weighted ambient temperature minus set temperature) increases and when containers are exposed to out of designed temperatures for long periods of time. Extreme temperatures tend to trigger continuous high-level power-output, which is another cause for temperature deviations. It is also important that the containers maintain a good battery level during shipments to avoid temperature deviations. For essential parts failure alarms, charger alarms are more likely to occur when the supply voltage is between 100V-127V. In this report, we also identified top 3 airlines and trade lanes with high temperature deviation rates to identify major improper container handling behaviors. According to our analysis, airlines should maintain an average battery level above 70%, avoid number of doors opening more than 5 times and minimize exposure to out of design temperatures. Artificial Neural Network with cost-sensitive learning was chosen as the best predictive model with a good validation AUC score of 0.936 and the best validation recall score of 0.891. The result after implementing the chosen model on the test set also yielded an AUC score of 0.936 and a recall score of 0.891.

Table of Contents

List of Figures

List of Tables

Introduction

Problem/Opportunity Statement

Each shipped cold chain container is highly valued, and every failure can be costly in the values of containers and people's life. Pharmaceutical manufacturers lose 15 billion per year due to cold chain failure (Cece, 2020). With over 75% of market share, Envirotainer is the biggest victim of cold chain failures. Temperature deviation triggered by manual error and unpredictable conditions during shipment is the No. 1 cause of cold chain failures. After multiple in-depth discussions with Envirotainer and a thorough examination of the sample data, the following issues need to be addressed through our data analysis:

- Many of the past failures were caused by improper container handling during the shipment, i.e., not charging the container or switching batteries within the required periodicity, allowing containers exposed to out of design ambient temperature for long periods of time, and opening doors too frequently during a shipment. By understanding the impact of improper container handling behaviors, Envirotainer could standardize and improve the training process for their clients and airlines to lower improper handling related failure rate.
- Containers tend to be more likely to fail in trade lanes that are exposed to extreme temperatures and have long shipment durations. It is mainly caused by the fact that airlines and shippers fail to provide temperature-controlled storage when containers are exposed to extreme ambient temperatures.
- The company's existing model detecting failures uses only the container performance data based on a manually set criterion. No risk profile was developed to characterize failure causes and no model was developed to predict failures.

- A lot of design and training hypothesis are not supported by real data analyses.

To address issues mentioned above, a risk profile will be built to identify top failure causes to reduce cost from failures. Failure analysis will be differentiated by failure types, i.e. temperature deviations, performance anomalies, and container parts failure alarms. Based on the risk profile, top failure causes will be categorized into the external uncontrollable causes and container handling behavior causes in order to provide appropriate recommendations. Most importantly, a model that predicts likelihood of shipment failure and pre-evaluates shipment performance before each shipment, will be developed; so that Envirotainer can implement preventative measures for shipments with high failure likelihood predictions.

Research Purpose

To identify container handling behaviors and airline trade lanes with extreme conditions that cause high failure rate as well as accommodating container production and design issues, following research objectives will be targeted to guide our analysis:

Failure Risk Profile

First, the profile will determine the impact of season (time of the year) and ambient temperatures on failures. It is also important to see the impact of the container battery level and charging condition during a shipment on failures. Furthermore, this risk profile will seek to learn the impact of shipment duration and number of door openings on failures. After a thorough understanding of the impact from above factors on failures, ranking airlines by the number of failures can help to identify the leading failure causing container handling behaviors. Lastly, the profile will also identify the trade lanes with the most frequent failures; the objective is to understand how trade lane external conditions impact failure rates.

Predictive Model

Given certain characteristics of a shipment order, a machine learning model will be developed to predict the likelihood of shipment failure. Answers to this question can prompt our client to keep a closer track and place more measures to ensure the success of a more likely to failure shipment.

Variables and Scope

Before the analysis, it is important to produce a clear and intuitive definition for failure for the failure risk profile. After in depth discussion, failures were defined and aggregated into a binary variable with 1 being failure and 0 being normal. 1 is assigned when one or more of the following failure conditions occur:

1. The average inside temperature of a container during a shipment deviates from the set temperature range.
2. One or more of the essential parts failure alarms, i.e. charger, compressor, and heater, were triggered during a shipment.
3. Container compressor performance anomalies (Containers overperformed or underperformed during a shipment). Overperformance and underperformance are identified using container power-steps vs. difference between ambient and set temperature.

The failure variable will be categorized into different types, i.e., temperature deviation, different alarm failures, and compressor performance anomalies. **Different failure types will be analyzed separately against relevant factors to build a failure profile for each failure type**.

To avoid average inside temperature values being over influenced by container's preconditioning period, only shipments with operational log time duration larger than 8 hours are chosen.

Background

Introduction to the Envirotainer Company

Envirotainer is a global market leader in pharmaceutical cold chain logistics that deliver temperature-sensitive pharmaceuticals. They are voted as the Best Active Temperature-Controlled Packaging Solution Provider at the Asia-Pacific Bioprocessing Excellence Award in 2018 (Markets Insider, 2018). As the leading company in the industry, Envirotainer plays an important role during this Covid-19 pandemic, and they are responsible for shipping Covid-19 vaccine for clinical trials (STAT Times, 2020). It is important for Envirotainer to understand and monitor the whole process of delivery because improper container handling or long exposure to extreme ambient temperatures will cause waste of the life-saving pharmaceuticals and will drive up the cost.

Pharmaceutical Cold Chain Logistic Industry Overview

The global pharmaceutical logistics market size was valued at USD 69.0 billion in 2019 and is expected to grow by 7.3% from 2020 to 2027 at a compound annual growth rate (Grand

View Research, 2020). North America and Europe are responsible for the lion's share of this global revenue, while Asia is catching up fast (VRR, 2020).

Pharmaceuticals are highly valued and sensitive to temperature. Pharmaceutical manufacturers lose 15 billion per year due to cold chain failure (Cece, 2020). Therefore, pharmaceutical logistics needs to optimize effectiveness rather than efficiency. Since many companies currently do not use temperature and location tracking containers, this indicates that companies do not pay much attention to container performance. Thus, companies still need to improve their management process and container qualification to reduce the risk of failure.

Literature Review

Based on the understanding of sample data, this failure detection problem was reframed as a classification problem with an imbalanced data set. A key challenge for this problem is that the classes are imbalanced. This makes the optimization of classification accuracy meaningless. To tackle this challenge, two approaches are considered. The first approach is cost-sensitive learning, and the second approach is adaptive re-sampling.

Cost-sensitive Learning

In cost-sensitive learning, off-the-shelf classification algorithms are used in the case of cost-sensitive learning. The only difference is that, in cost-sensitive learning, the loss functions of classification algorithms are modified to weight the classification errors differently for normal instances and failure instances. In this case, the misclassified cost is considered in the loss function. For example, in binary classification, cross-entropy is usually used as the loss function:

$$Loss(\hat{y}_i, y_i) = \sum_{i=1}^{n} (-y_i log(\hat{y}_i) - (1 - y_i) \; log \; log \; (1 - \hat{y}_i) \;)$$

However, cross-entropy is not designed for imbalanced data sets. In an imbalanced data set, only a small number of instances are minority classes. If those minority classes are entirely misclassified, the value of the loss function can still be very small and this small loss function value is meaningless in practice. Therefore, loss function should be modified with different weight coefficients applied to different classes to take the imbalanced class sizes into consideration. Please see the equation below:

$$Weighted \; Loss(\hat{y}_i, y_i) = \sum_{i=1}^{n} (-w_0 y_i log(\hat{y}_i) - w_1(1 - y_i) \; log \; log \; (1 - \hat{y}_i) \;)$$

The loss functions of many classification algorithms can be modified to account for the misclassification cost. These classification algorithms include Logistic Regression, Random Forest Classifier, Support Vector Machine and Deep Learning, etc.

Adaptive Re-sampling

The concept of adaptive re-sampling is first mentioned by Japkowicz in her paper on random sampling (Japkowicz, 2000). Japkowicz discussed the effect of imbalance in a dataset and mentioned two resampling approaches. One approach was over-sampling the minority class at random with replacement until it consisted of as many samples as the majority class. Another approach was under-sampling the majority class at random until it reached the sample size of the minority class. In her paper, Japkowicz noted that both resampling approaches were effective.

However, there is a problem associated with over-sampling minority class with replacement. In over-sampling minority class with replacement, minority class was replaced repeatedly, creating many duplicates. In this case, when use classifiers, such as decision tree classifier, more nodes will be created as the algorithm tries hard to learn more about those duplicates. This may lead to overfitting. That's where SMOTE comes in.

SMOTE is short for Synthetic Minority Over-sampling Technique (Nitesh V.Charwla, et al., 2002). Rather than over-sampling minority class with replacement, SMOTE over-samples the minority class by creating "synthetic" examples. Those "synthetic" examples are generated by randomly taking points along the lines between each minority class sample and its k minority class nearest neighbors. Nitesh V.Charwla and his colleagues did a bunch of experiments to prove that SMOTE is superior to the ordinary over-sampling approach (oversampling minority class with replacement). Apart from that, their experiments also proved that combining the SMOTE with under-sampling majority class created even better results. Therefore, the team plans to adopt this approach to the analysis. Once a balanced data set is generated, the off-the-shelf classification algorithms can be applied.

Copula Based Anomaly Detection

To evaluate the performance of compressor, copula was utilized to detect the performance anomalies. The concept of copula is to separate the marginal distribution from the dependency information of given random variables. In other words, copula is the joint distribution of uniformly distributed random variables.

$$Copula = F_{U_1, U_2, \dots U_n}(u_1, u_2, \dots, u_n)$$

where $F(x_1, x_2, \ldots, x_n)$ is the joint distribution function.

Let random variables $X_1, X_2 \ldots, X_n \sim F_{X_n}$, where F_{X_n} are the marginal distributions. To remove the influence of marginal distribution from the joint distribution, it is suggested to first transform the random variables into uniform marginals so that later they can be transformed into any desired distributions. $(U_1, U_2 \ldots\ U_n) = (F_1(X_1), F_2(X_2), \ldots, F_n(X_n))$, where$(U_1, U_2 \ldots\ U_n) \sim Uniform(0,1)$. Once they are transformed into uniform marginals, their copula function can be found by analyzing dependence patterns.

In the copula world, Spearman rank correlation coefficient (ρ_S) or Kendall rank correlation coefficient (ρ_τ) are used to measure the monotonic dependence.

$$\rho_S(X, Y) = \rho(F_X(X), F_Y(Y))$$

$$\rho_\tau(X, Y) = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$$

If ρ_S or ρ_τ = 1 then X and Y are comonotonic; if ρ_S or ρ_τ = -1 then X and Y are counter-monotonic.

To detect anomalies, quantile regressions are used to calculate the quantiles of level α of conditional distributions of copulas. If the copulas show lower tail dependence (Clayton Copula), the formula below is utilized.

$$C_{2|1}^{-1}\{u\} = \left(\left(\alpha^{\frac{-\theta}{1+\theta}} - 1\right) u^{-\theta} + 1\right)^{-\frac{1}{\theta}}$$

Methodology

Analysis Approach

Analytics workflow started with descriptive data analysis, through which individual failure risk profiles were developed. These failure risk profiles can help to identify the top causes of each failure type and production and design issues with real data analytics. The airline container handling patterns can give insights on how airlines handle containers during a shipment, which can help to discover the handling behaviors that lead to high failures ratio and give training suggestions to improve their container handling. Based on the descriptive analytical profiles, a predictive analysis was also conducted. Given certain information for a particular shipment, i.e., trade lane, airline, season, likely ambient temperature the container will experience during the trip, etc., a likelihood of failure can be predicted using supervised machine learning classification. This predictive analysis can help the company to allocate more resources, give pin-pointed handling instructions, or reconsider flight routes for a shipment when a prediction of high likelihood of failure is given. It is expected the data to be very imbalanced, meaning the vast majority of the shipments will not be failures. In order to obtain a meaningful result, techniques like adaptive-resampling and cost sensitive learning were applied to address the imbalanced data problem. The performance anomalies and failure alarms are innate failures of the containers, but the temperature deviation is a derived variable based on the average inside container temperature during a shipment. To validate project team’s definition of temperature deviation, a small set of reported temperature deviation incidents were used to show the similar trend of self-defined temperature deviation and reported temperature deviation against different factors.

Data

Three data sets were provided to the team by Envirotainer, container logged data, trade lane & shipment data as well as the reported incidents data. The container logged data set consists of variables that were logged by container sensors during a shipment (See Table I). The trade lane & shipment data set is comprised of variables that contain logistic information of each shipment (See Table I). The reported incidents data contains information regarding the actually reported incidents (See Table I). The three data sets are integrated using container number (unique for each container) and order number (unique for each shipment). Each observation contains information regarding a unique container movement from origin to destination. It is important to note that there could be multiple containers in one shipment and failures may only occur in one or some of the containers of the same shipment. There are two type of containers Envirotainer provides: E-container, an advanced temperature-controlled air cargo container that utilizes compressor cooling and electric heating technology as well as T-container, a cost-effective container that utilizes dry-ice cooling to control temperature. **Each container type has a unique set of logged parameters and analyses mentioned in the approach section were only applied to E-containers as the T-containers being phased out of the service**. Three years of data (Jan 2018- Feb 2021) were provided to the team with an average of more than 25,000 unique container movements for each year. Additional supplementary data, airport countries' supply voltage and economic status information, were researched and organized into a supplementary data set. (DataHub 2019, DataHub 2020, Wiki 2021, and The United Nations 2014)

Three variables were aggregated from the raw data for analytic purposes, the Out of Design Temperature Exposure Duration (ODTD), the Average Weighted Ambient Temperature (AWAT) and Delta. ODTD is the amount of time a container was exposed to out of design extreme ambient temperatures during a shipment. AWAT is the averaged ambient temperature weighted by the duration a container was exposed at each temperature level during a shipment. Delta is the difference between AWAT and set temperature during a shipment.

Other container performance and logistic variables along with the three aggregated variables mentioned above were used as independent features for both descriptive and predictive analyses. (See Table 1)

Table 1. *Definitions of different variables in Container Logged Data, Trade Lanes & Shipment Data, Reported Incidents Data and Supplementary Data*

Data Sets	Features	Meta Data
Container Logged Data	Trip Reference number	Unique to each container movement
	Container Temperature	Mean Temperature inside each container during a shipment
	Different Ambient Temperature Range Duration	Amount of time the container experiences at different ambient temperature range
	Set Temperature	Desired inside temperature set by clients
	Set Temperature limit	The allowed deviation limit based on set temperature (+/-3 ^{0}C)
	Amount of time a container battery level was below 30%	E-Containers only
	Mean Battery level During a Shipment	E-Containers only
	Number of Door Openings	
	Battery Charging Duration	E-Containers only

	Compressor Failure Alarm	E-Containers only
	Heater Failure Alarm	E-Containers only
	Charger Failure Alarm	E-Containers only
	Mean Power Steps (Power output level) during a Shipment for both cooling and heating	E-Containers only
Trade Lanes & Shipment Data	Shipment Duration	
	Airline Name	Main shipment handler
	Container Number	Unique for each container
	Order Number	Unique for each shipment
	Trip Reference number	Unique to each container movement
	Shipment Origin	
	Shipment Destination	
	Shipment Trade lane	
	Shipment Release Date	
	Container Type	
Reported Incidents Data	Container Number	Unique for each container
	Order Number	Unique for each shipment
	Trip Reference number	Unique to each container movement
	Case Origin	How was the incident discovered?
	Incident Category	Type of Incident
	Cause Category	Cause of the Incident
	Case Category	Solution Type
Supplementary Data	Airport Code	
	Country Name	
	Country Code	
	Supply Voltage	
	Economic Status	Developed or Developing Country

Quantitative Methods

The methodology proposed for the analysis consists of the following sections.

Exploratory Data Analysis and Imbalanced Data Handling

Through data exploration, missing values will be imputed using interpolative or regressive imputation. Bivariate correlation will be analyzed to address collinearities among different features. Different techniques of dealing with imbalanced data, i.e., SMOTE (Blagus & Lusa, 2013), random under-sampling and cost-sensitive learning, are explored to identify the best strategy to address the imbalanced failure label class sizes.

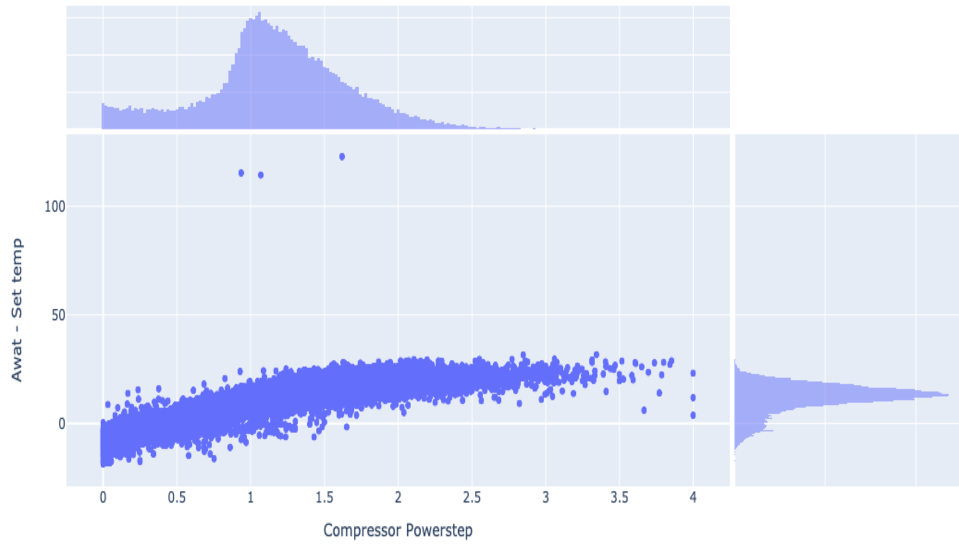
Compressor Performance Anomalies

Compressor is most utilized part of the containers to maintain the inside temperature. Therefore, identifying the root causes of compressor performance anomalies is essential to our failure analysis. Analyzing how compressor power step works when the Delta changes can help to evaluate the performance of compressor power step.

$$Delta = Average\ Weighted\ Ambient\ Temperature - Set\ Temperature$$

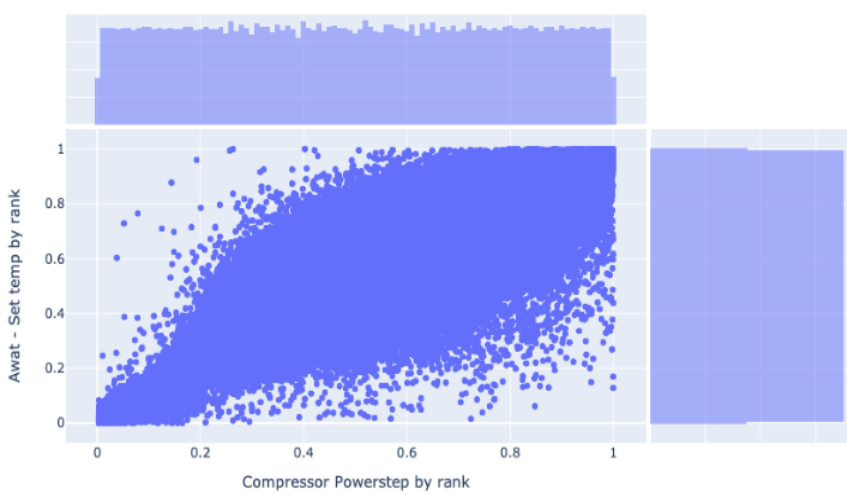
According to Figure 1, the marginal distributions of both compressor power step and Delta are not normal.

Figure 1. *The Original Distribution of Compressor Power Steps and Delta*



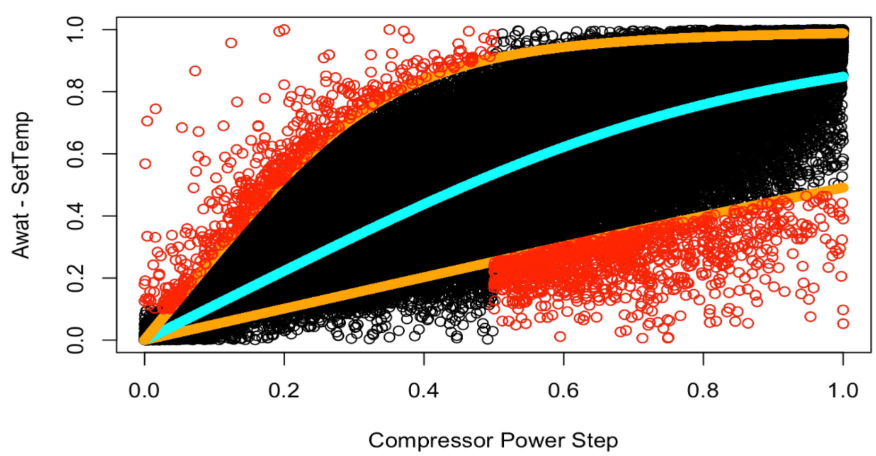
Therefore, copula can be used to find the dependency between compressor power step and Delta, in which the marginal distributions of both compressor power step and Delta are transformed into uniform distribution. To transform the marginal distributions of these two random variables into uniformly distributed, all the data points are ranked and normalized by the total number of the data points (Figure 2).

Figure 2. *The Transformed Distribution of Compressor Power Steps and Delta*



Once the random variables are transformed into uniform distribution, Clayton copula and quantile regression are used to detect anomaly data points. Anomaly is defined as the deviation from comonotonic dependency between power step and Delta with 5% and 95% quantiles (Figure 3). The upper-left data points are identified as under-performing compressor power step while the bottom-right data points are over-performing compressor power step (Figure 3). Finally, all anomaly data points were extracted for further analysis.

Figure 3. *Clayton Copula 5% and 95% Quantiles*



Failure Risk Descriptive Analyses

With the objective of gaining insights on relevant causes of each failure type, individual failure risk profiles are generated. For each failure type, the frequency of failures for each airline was counted to identify airlines with highest rate of failures, subsequently the container handling behaviors, e.g. charging, ambient temperature control and door openings, that lead to a high failure rate were identified. More importantly, the impact of other trade lane and shipment related factors like, shipment duration, season (time of the year), ODTD, AWAT, etc., on failures were analyzed to identify the top failure causes that are not due to human error.

Supervised Machine Learning Classification and Prediction

The ultimate objective of finding root causes for different failure types is to avoid temperature deviation during a shipment. Thus, temperature deviation was chosen as the label for the predictive analysis. With temperature deviation as the label and other relevant features that were given or can be estimated at the beginning of a shipment as predictors, different machine learning models (logistic regression, XGBoost, and Artificial Neural Network) together with performance-boosting methods (adaptive resampling and cost-sensitive learning) were utilized to come up with the best model to predict the likelihood of temperature deviation during a trip. Before a shipment starts, given the information of the shipment, the model is able to give a fairly accurate prediction on the likelihood of temperature deviation.

Only features whose values are given or can be estimated at the beginning of each shipment were chosen as predictors, which were the time of the year, set temperature, weighted average ambient temperature, log time, usage count and usage time (Shipment numbers and total previous shipment duration a container logged at the beginning of the shipment), origin-destination continent, origin-destination country status, and airlines. The last three features were recategorized from the information of trade lanes and airlines since both trade lanes and airlines contain too many levels if used as categorical features. The information of trade lanes was separated into two features: one records the continent of the origin and destination, and the other records the economic status of origin and destination (whether origin or destination is a developed country or not). For airline categories, airlines are categorized into three groups: airlines with more than 2,000 Envirotainer shipments are defined as “big airlines”, while with 500 ~ 2,000 Envirotainer shipments are “medium airlines”, and the rest are “small airlines”.

Model Evaluation

The data was split into train (67.5%), validation (22.5%), and test (10%) sets for predictive supervised machine learning classification. Different machine learning models were trained on the train set and then fine-tuned on the validation set. The machine learning model that does not overfit and scores highest on the validation set was picked as the best predictive model and will be further implemented on the test set.

Due to the fact that failures only occur in a very small fraction of shipments and are very costly to pharmaceutical companies as well as human life, it is preferred to over predict failures and place extra caution and preventative measures during a likely to fail shipment. With the above reasoning in mind, area under the curve (AUC) scores and recall score were used for model evaluation.

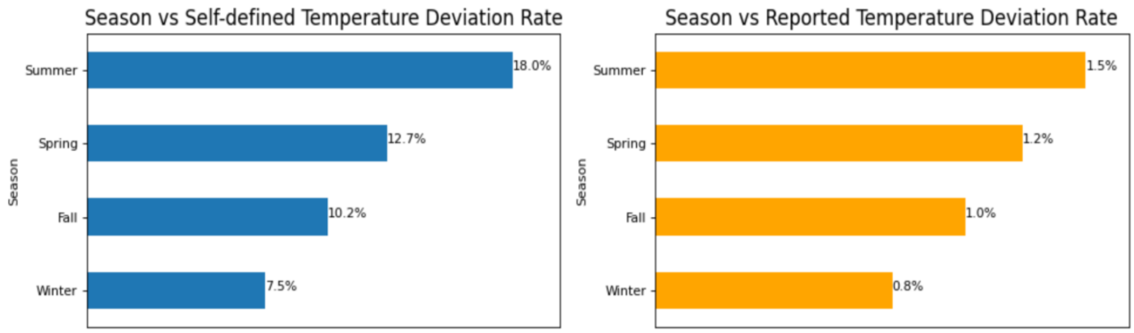
Results and Findings

Temperature Deviation

A comparison analysis was conducted between reported temperature deviations (RTD) and self-defined temperature deviations (TD). This comparison analysis aims to validate the definition of temperature deviation. Due to the large difference in size (12.3% of the data are TDs and 1.1% of the data are RTDs), only non-human related factors, e.g., container workloads and shipment environment related factors, are used for the comparison.

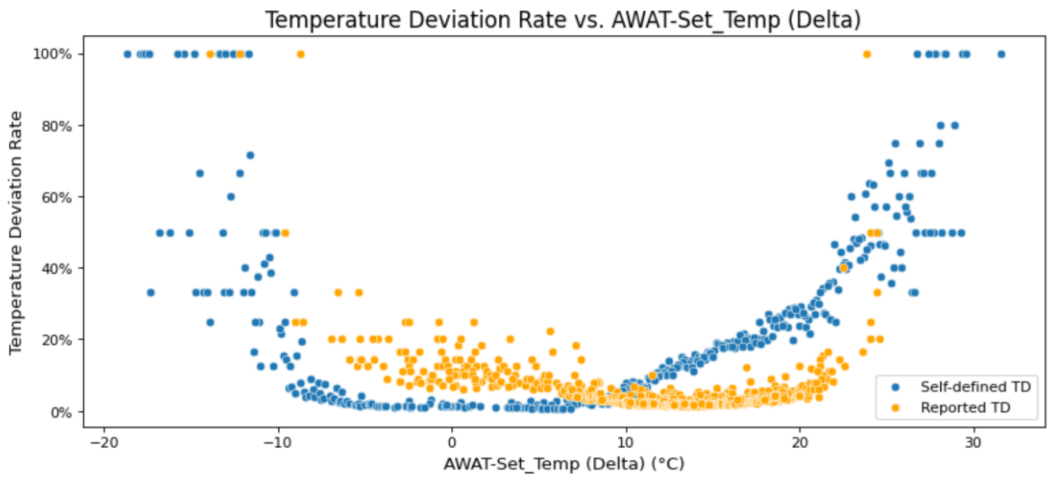
Since pharmaceuticals are temperature-sensitive, temperature deviation often happens when the weather temperature is relatively high. We can see from the left side of Figure 4 that the temperature deviation rate is highest in summer, between June and August. Based on the right side of Figure 4, the reported temperature deviations show the same trend against the season factor.

Figure 4. *Self-defined Temperature Deviation Rate vs. Season (Left)*
Reported Temperature Deviation Rate vs. Season (Right)



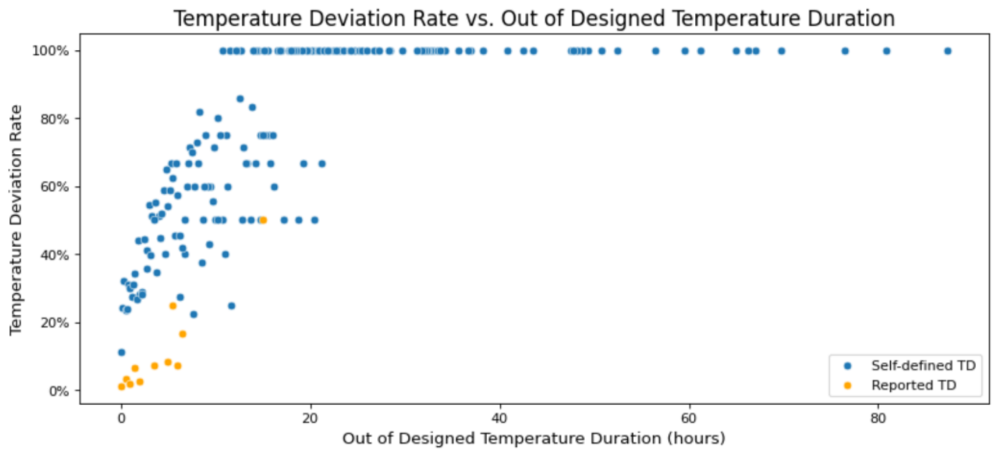
Considering temperature-sensitivity, the temperature deviation rate increases as the difference (Delta) between set temperature and average weighted ambient temperature increases. The normal shipments' Delta mean is 10.95°C, while the temperature deviation shipments' Delta mean is 15.69°C. We can see from Figure 5, when the Delta is extremely high or low, the temperature deviation rate is 100%. The temperature deviation rate minimizes when Delta is between -10°C and 10°C and starts to exponentially increase when Delta is over 10°C or below -10°C. The reported temperature deviation rate minimizes when Delta is between 0°C to 15°C and exponentially increases when Delta is over 20°C and below -10 °C. The trend of self-defined temperature deviation is very similar to the trend of reported temperature deviation.

Figure 5. *Temperature Deviation Rate vs. Average Weighted Ambient Temperature - Set Temperature (Delta)*



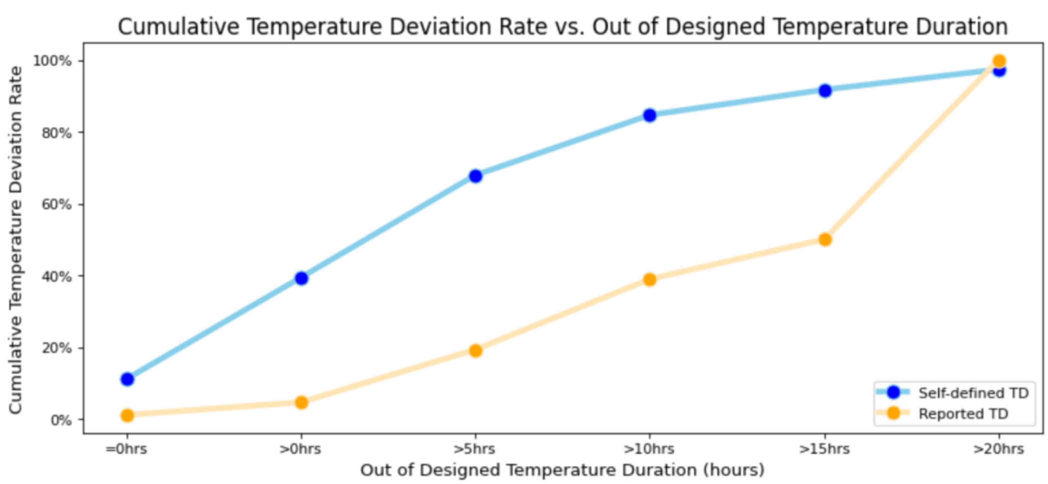
There is a desired range of ambient temperature the containers are designed to operate under. The containers are expected to underperform when facing out of designed ambient temperature. The analysis shows that the duration of containers exposed to out of designed temperature matters to temperature deviation. The out of designed temperature duration (ODTD) mean for non-temperature deviation shipment is 0.067 hours, while the temperature deviation shipments' ODTD mean is 0.83 hours. Likewise, when a container exposes to longer ODTD, it is more likely to cause temperature deviation. The reported temperature deviation shows the similar trend. (See Figure 6)

Figure 6. *Temperature Deviation Rate vs. Out of Designed Temperature Duration*



As we take a closer look at the cumulative TD rate vs. ODTD (Figure 7), when the ODTD increases from 0 hour to 10 hours, the cumulative temperature deviation rate significantly increases from 0.11 to 0.87. The reported temperature deviations show a similar cumulative TD rate trend against ODTD. As ODTD increases, reported Temperature Deviation rate also increases.

Figure 7. *Cumulative Temperature Deviation Rate vs. Out of Designed Temperature Duration*



Shipment duration is also a factor that affects temperature deviation. The mean shipment duration for non-temperature deviation shipments is 130.63 hours, while for temperature deviated shipments is 233.15 hours. Temperature deviation rate increases from 0.01 to 0.08 when the shipment duration changes from less than 5 days to less than 10 days (Figure 8). The reported temperature deviations show a similar cumulative TD rate trend against shipment duration. (See Figure 9)

Figure 8. *Cumulative Self-defined Temperature Deviation Rate vs. Shipment Duration*

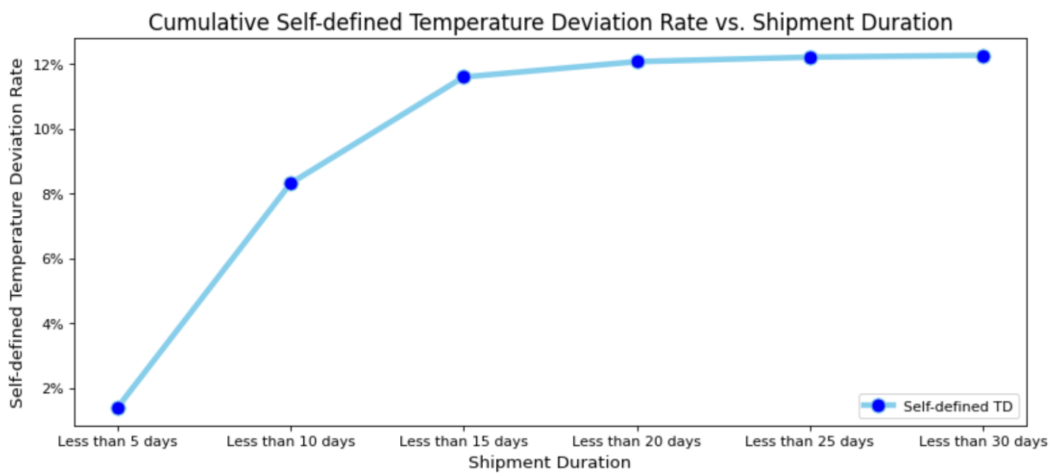
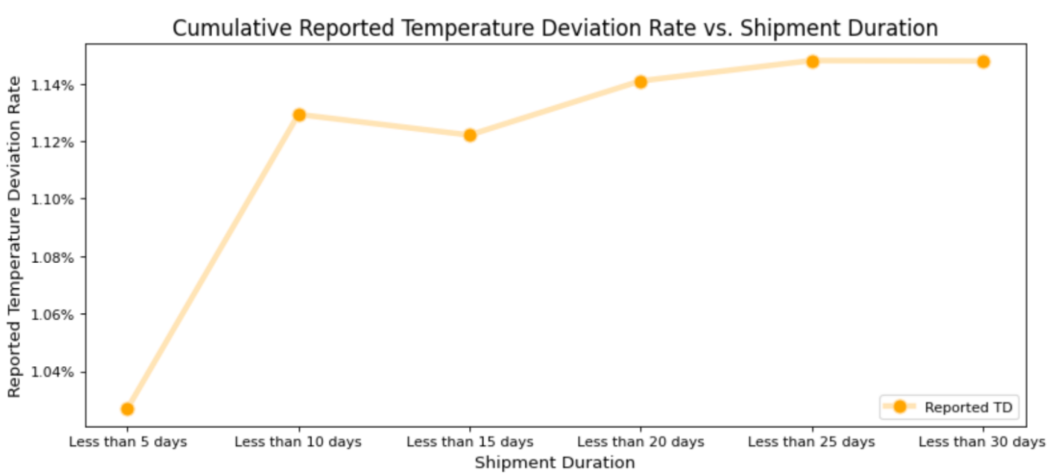


Figure 9. *Cumulative Reported Temperature Deviation Rate vs. Shipment Duration*



Other than the factors that containers experienced during the shipment, we also want to examine if the origin and destination country place any effect on container temperature deviation. In this case, we did a Chi-squared test since countries are categorical variables, and we defined the origin country and destination country as either developed or developing country. Four categories of country economic status are developed to address this problem:

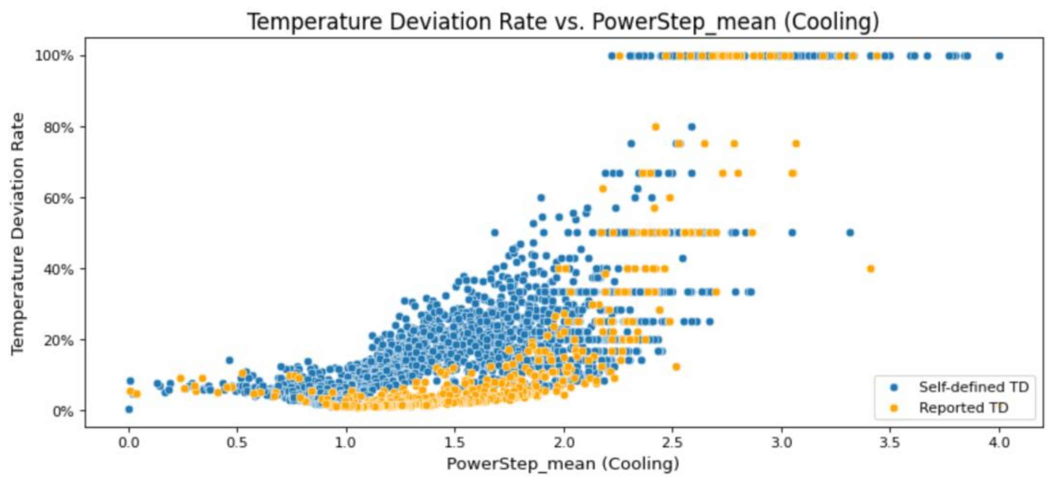
- Developed country to developed country
- Developed country to developing country
- Developing country to developed country
- Developing country to developing country

The null hypothesis is that all four categories of economic status of origin and destination country have the same ratio of temperature deviation, while the alternative hypothesis is at least one of these combinations does not have the same ratio of temperature deviation.

The result of our Chi-square test reached 0.05 statistically significance with a p-value of 0. Conclusively, the economic status of a country does affect the shipment temperature deviation rate. The temperature deviation rate is highest at 17.19% when both the origin and destination are developing countries. Economic status of origin countries has a higher impact on temperature deviation ratio as the airline that handles the shipment tend to come from the origin country. Developed country airlines tend to handle containers better and produce smaller TD ratios.

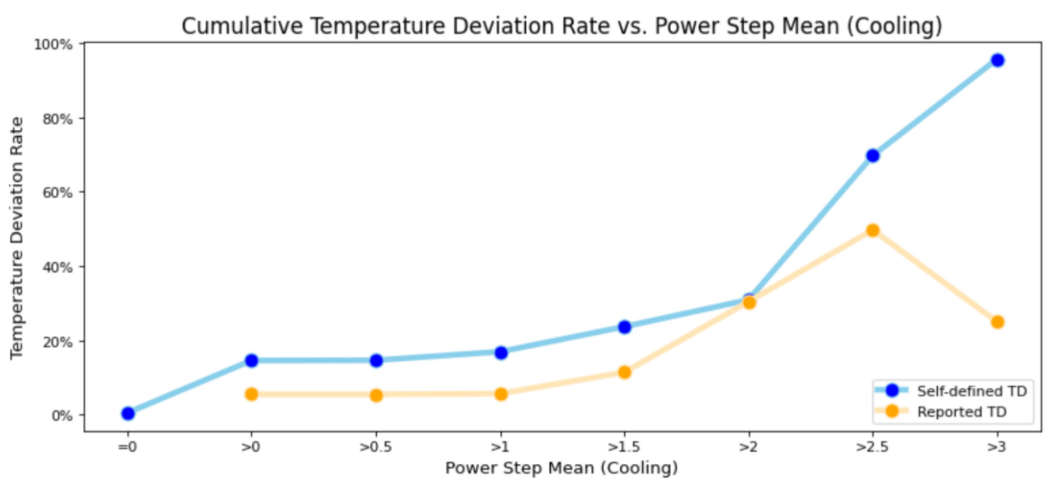
We are also interested in how container workload affects temperature deviation. Figure 10 shows that as the power step used for compressor (workload) increases, the containers experience higher temperature deviation rate. The reported temperature deviation rate demonstrates a very similar trend as its self-defined counterpart when plotted against mean compressor power step.

Figure 10. *Temperature Deviation Rate vs. Power Step Mean (Cooling)*



There is also a marked increase in cumulative temperature deviation rate when the mean compressor power step is getting larger. The cumulative reported temperature deviation rate follows a similar trend. (See Figure 11)

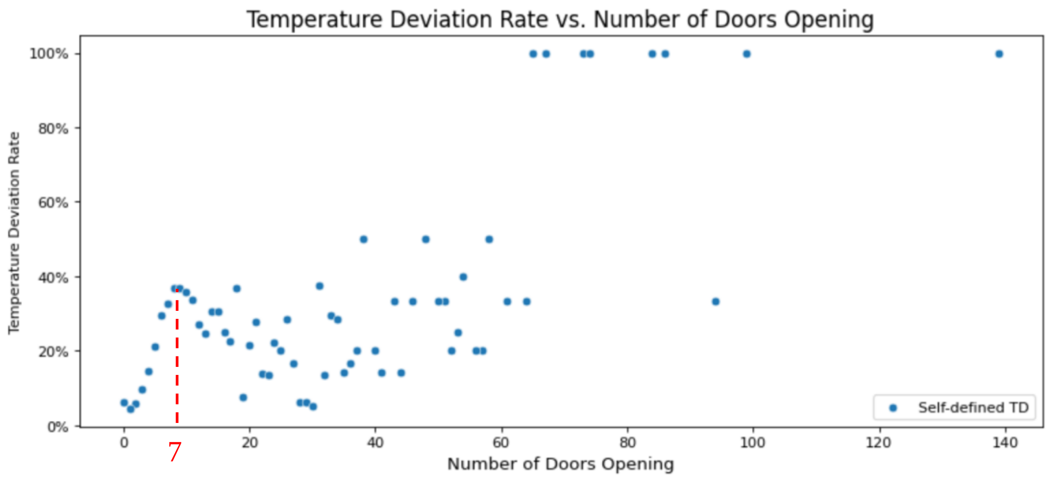
Figure 11. *Cumulative Temperature Deviation Rate vs. Power Step Mean (Cooling)*



Next, factors that are related to container handling errors, such as the number of door openings during the shipment and battery conditions, are examined.

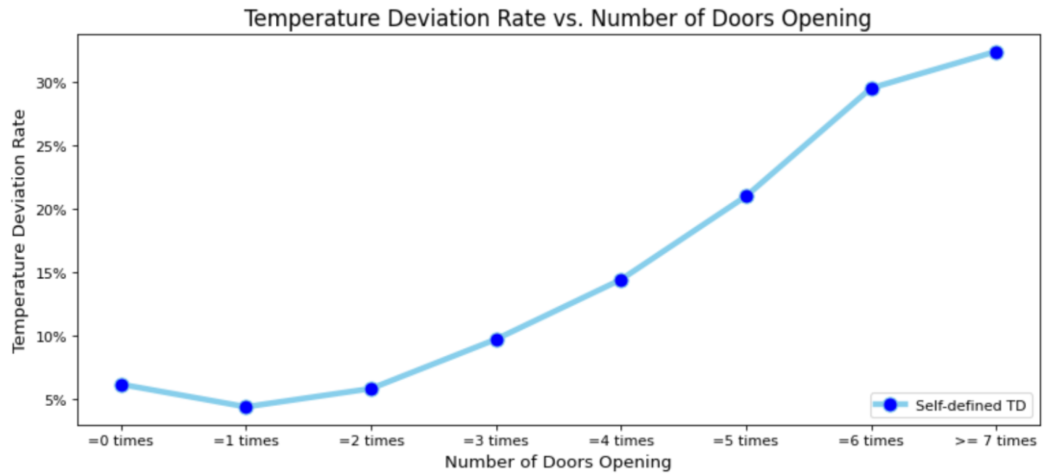
The number of door openings during a shipment is more likely to cause temperature deviation as this number increases. The number of door openings mean for non-temperature deviation shipment is 3.34 times, while for the temperature deviated shipment is 4.97 times. Since the majority of the data have less than 7 times of door openings, we should group data with more than 7 times of door openings into one category. The increasing trend of TD Rate vs. Number of door openings is clearly indicated in Figure 12 when door opening is less than 7.

Figure 12. *Temperature Deviation Rate vs. Number of Doors Opening*



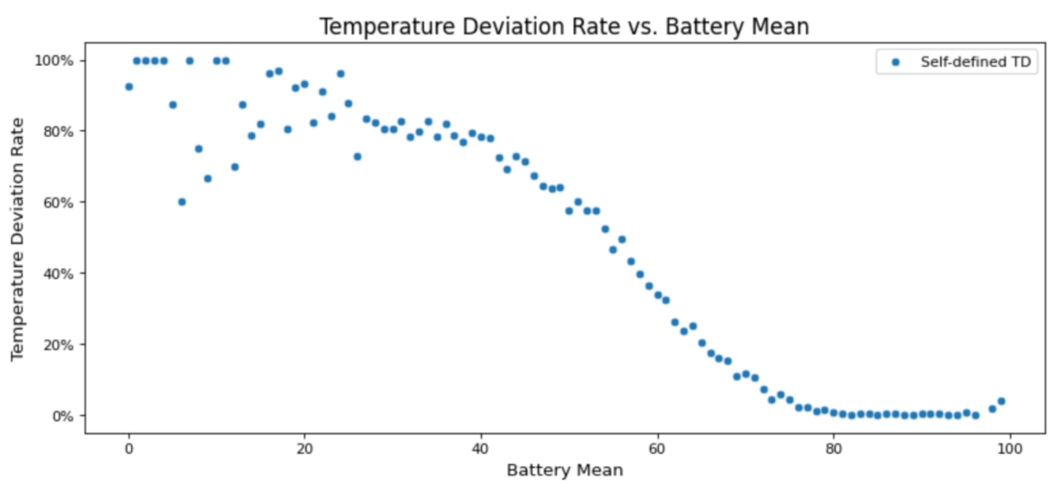
When we take a closer look at the number of door openings between 0 to 7 times, the temperature deviation rate increases from 0.06 to 0.32. The temperature deviation rate climbs significantly when the number of door openings is above 3 times during a shipment. (See Figure 13)

Figure 13. *Temperature Deviation Rate vs. Number of Doors Opening (Grouped)*



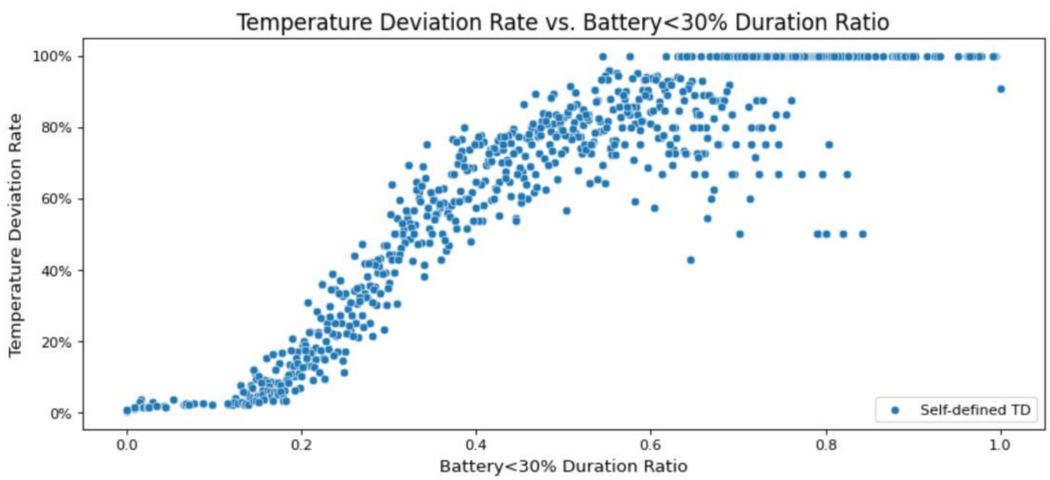
Additionally, the mean battery level during a shipment generates a noticeable impact on temperature deviation. The non-temperature deviation shipments' battery mean is 83.82%, and temperature deviated shipments' battery mean is 46.19%. According to the Figure 14, temperature deviation rate decreases as the battery mean increases. Continuously maintaining the battery above 70% is vital for reducing the TD rate.

Figure 14. *Temperature Deviation Rate vs. Battery Mean*



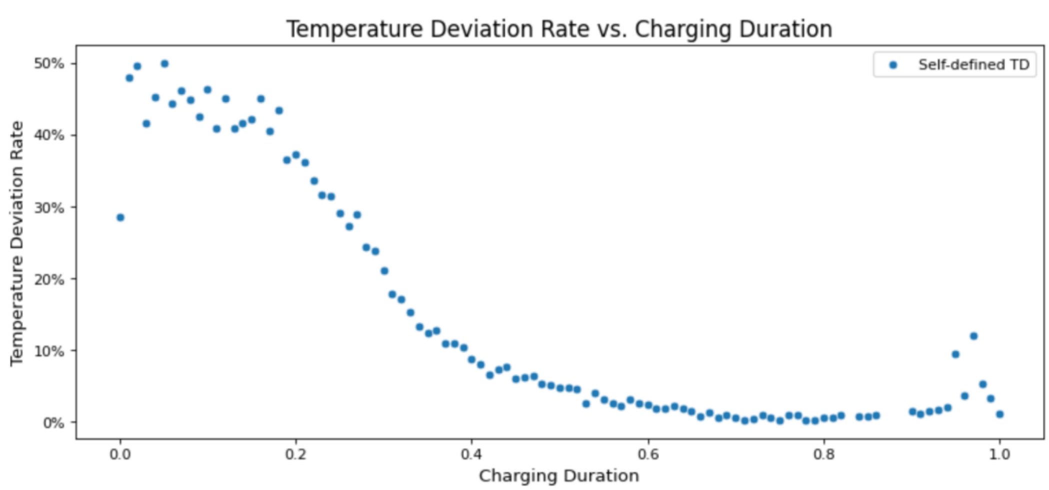
Also, it is suggested not to keep the container battery less than 30%. As the duration ratio of battery less than 30% increases, the temperature deviation rate will also increase. (See Figure 15)

Figure 15. *Temperature Deviation Rate vs. Battey<30% Duration Ratio*



Similarly, we could expect that the longer we charge containers during a shipment, the lower temperature deviation rate will be (See Figure 16).

Figure 16. *Temperature Deviation Rate vs. Charging Duration*



Airlines and Trade Lanes with High TD Rate

To help decrease the temperature deviation rate, Envirotainer should understand airline's container handling behavior. In the airline analysis, we identified the top 3 temperature deviation rate airlines with large enough total shipment count.

To analyze the airline's overall container handling behavior, we used the selected top 3 airlines combined with their top temperature deviated trade lanes. According to Table II, they have a lower battery mean than non-temperature deviated shipments. The battery less than 30% and 0% duration are also higher, while the charging duration rate is lower than non-temperature deviation shipments. The average number of door openings is higher for those airlines than the non-temperature deviation shipments. Considering the above results, Envirotainer should pay more attention to how well the airlines maintain the container battery level during a shipment, how effectively airlines avoid out of design temperature exposure and how many times airlines open the containers during a shipment.

Table 2. *Top 3 Temperature deviation rate airlines' container handling behavior*

	Battery mean	Battery < 30 duration ratio	Battery = 0 duration ratio	Charging duration ratio	Number of door openings mean
Non-Temp Deviation	83.821543	0.043385	0.020093	0.498377	3.342965
Temp Deviation	46.193959	0.452953	0.381411	0.252690	4.967364
Airline78	52.309369	0.380164	0.306721	0.244777	3.845466
Airline8	56.229293	0.335765	0.241829	0.279544	3.842679
Airline46	59.434520	0.295381	0.244963	0.296757	5.324763

Shipment trade lanes with the highest TD rate tend to have a much larger average Delta than trade lanes with low temperature deviation rate. Containers in those high TD rate trade lanes also experience much longer ODTD. (See Table 3)

Table 3. *Top 3 Temperature deviation rate trade lanes container ambient temperature information*

	Delta Mean	ODTD Ratio
Non-Temp deviation	10.946935	0.066960
Temp deviation	15.691380	0.834061
ORD_FCO	15.853284	0.089730
BSL_JFK	11.030490	0.375570
LUX_HSV	15.436840	0.500845

Parts Failure Alarms

The alarms analysis mainly focuses on three major alarms of E-container: charger alarms, compressor alarms, and heater alarms. Chi-squared tests and trend analysis are conducted on alarms data. In Chi-squared test, number of alarms is used directly, while in trend analysis, the alarm rate is used instead. Alarm rate is defined as the number of alarms records divided by the total records given a certain situation.

Charger Alarms

Envirotainer is especially interested in the voltage analysis against charger alarm since the customer service team usually receives incidents report on charger alarms from customers in Japan and other countries with a supply voltage between 100V-127V. The voltage data for different origin and destination countries has been collected and then categorized into four groups:

- Group 1 contains countries with origin voltage of 100V - 127V and destination voltage of 100V - 127V
- Group 2 contains countries with origin voltage of 100V - 127V and destination voltage of 220V - 240V

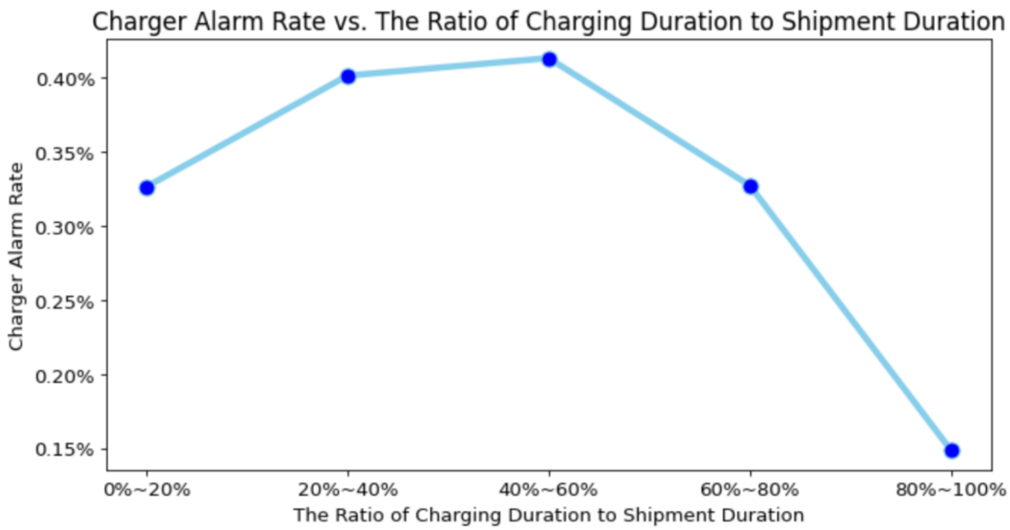
- Group 3 contains countries with origin voltage of 220V - 240V and destination voltage of 100V - 127V
- Group 4 contains countries with origin voltage of 220V - 240V and destination voltage of 220V - 240V

We examined whether the charging voltages place any effect on the ratio of charger alarms. Since the response variable is categorical variable, Chi-squared test was chosen. Before any test, we defined our null hypothesis as all four groups have the same ratio of charger alarms, while the alternative hypothesis states that at least one of these groups does not have the same ratio of charger alarms.

The result of Chi-squared test reached 0.05 statistically significance with a p-value of 0.003, which implies that voltage has an effect on the number of charger alarms. When the supply voltage of both origin and destination countries are between 100V and 127V, the charger alarm rate is the highest (0.65%). When the supply voltage of both origin and destination countries are between 220V and 240V, the charger alarm rate is the lowest (0.32%). The container chargers are twice as likely to fail when the supply voltage is between 100V-127V.

How does charging duration affect the charger alarm rate during a shipment? (See Figure 17). The general trend shows that the charger alarm rate will decrease if the containers are properly charged above 60% of its shipment duration. If the containers are charged above 80% of its shipment duration, there will be a sharp drop in the charger alarm rate. However, the charger alarm rate peaks if the container is charged for half of its total shipment time, which indicates that a frequent alteration between charging and not charging status will cause charger failure.

Figure 17. *Charger Alarm Rate vs. The Ratio of Charging Duration to Shipment Duration*



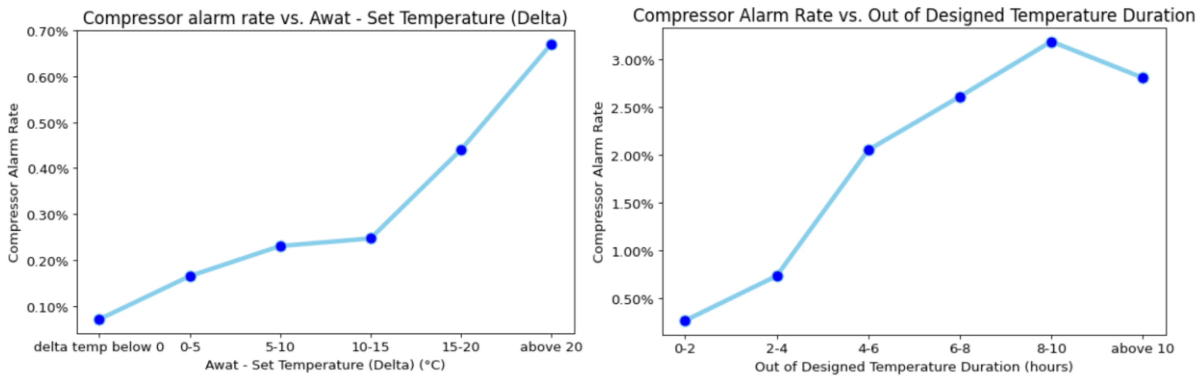
Compressor Alarms

The summer compressor alarm rate of 0.44% is significantly higher than the other seasons (Spring 0.21%, Autumn 0.25%, and Winter 0.28%), as confirmed by a chi-squared test with a p-value of 0.0002. The small p-value indicates that seasons pose an influence on the compressor alarm rate and summer has the highest compressor alarm rate.

The difference (Delta) between the weighted average ambient temperature (AWAT) and the set temperature is another critical factor affecting compressor alarms (See the left side of Figure 18). If Delta is above 0, the compressor will start to work in order to maintain its set temperature. According to the graph on the left-hand side, the compressor alarm rate goes up as the delta temperature climbs. Especially after the temperature reaches above 15°C, the compressor alarm rate climbs quickly.

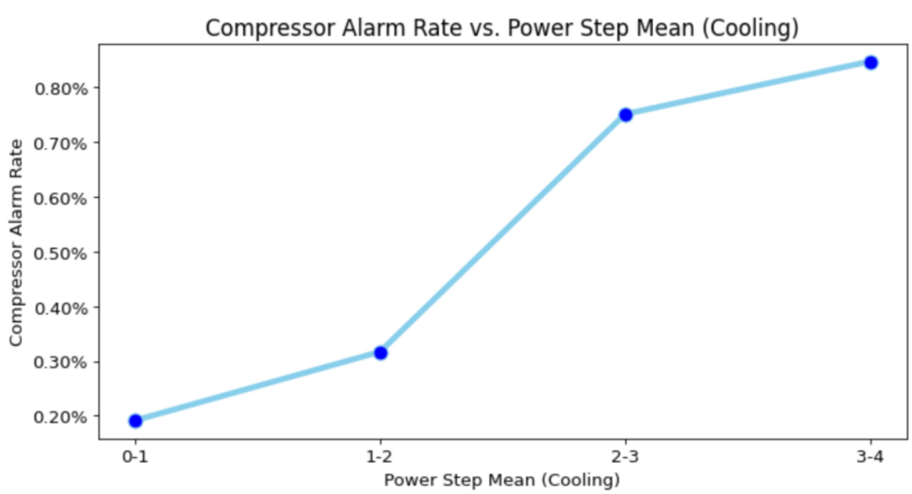
Out of designed temperature duration (ODTD) records how long a container experiences the extreme or out of design ambient temperature. As ODTD increases, compressor alarm rate increases as well. It seems that after total 4 hours of experiencing out of designed temperature, there is a big jump in the compressor alarm rate. (See the right side of Figure 18)

Figure 18. *Compressor Alarm Rate vs. Delta Temperature (Left)*
Compressor Alarm Rate vs. Out of Designed Temperature Duration (Right)



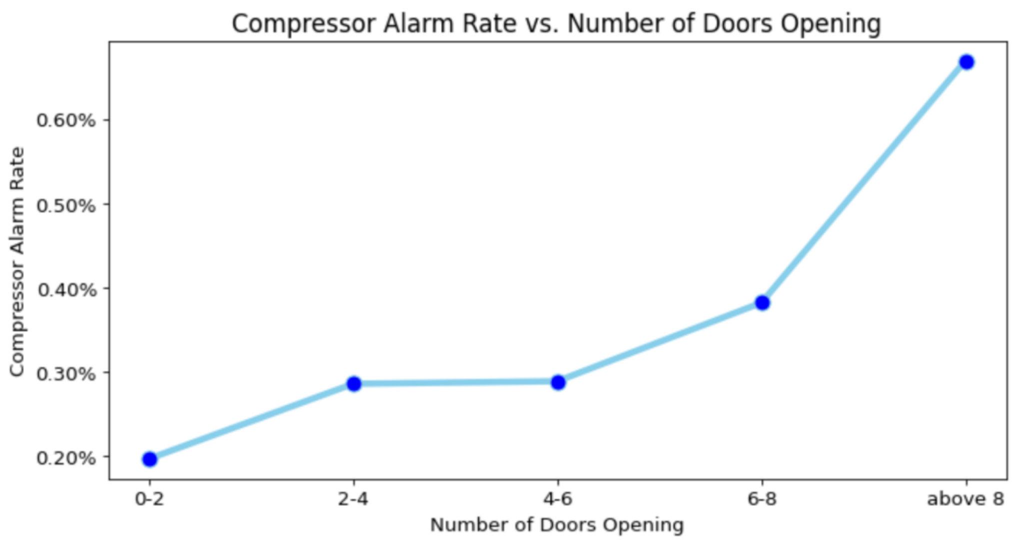
A higher power step mean suggests that the container is working harder to maintain its set temperature. According to our analysis result, the higher the power step mean is, the more likely the container is to trigger compressor failure alarms. A critical value of power step mean of 2, after which there is a big jump in the compressor alarm rate, is observed. (See Figure 19)

Figure 19. *Compressor Alarm Rate vs. Power Step Mean (Cooling)*



The compressor alarm rate also climbs as the number of door openings increases. Especially after the number of door openings reaches 6 time or more, there is a big jump in the compressor alarm rate. (See Figure 20)

Figure 20. *Compressor Alarms Rate vs. Number of Doors Opening*

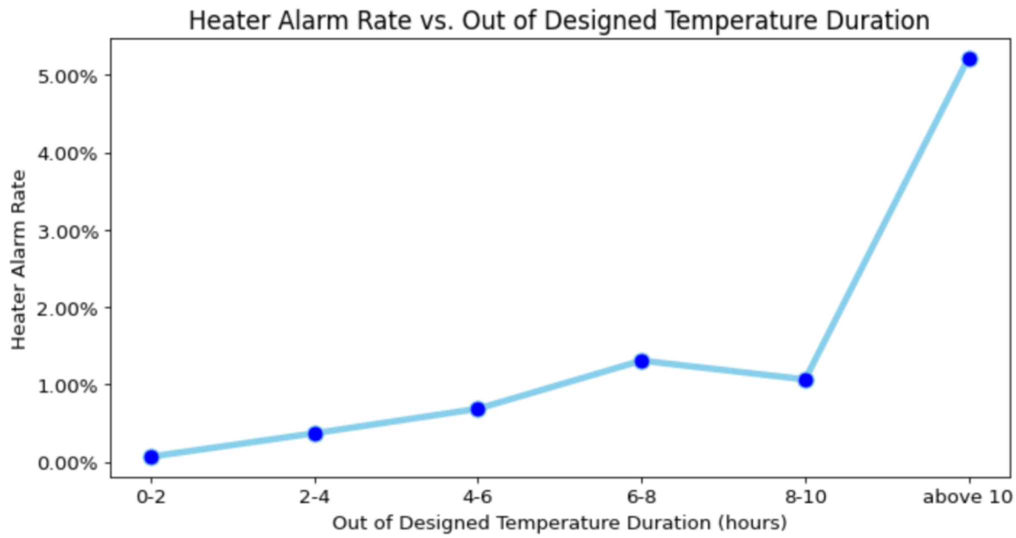


Heater Alarms

Similar to compressor alarms, heater alarms are more likely to occur in summer. The summer heater alarm rate of 0.20% is significantly higher than other seasons (Spring 0.06%, Autumn 0.04%, and Winter 0.06%), as confirmed by the chi-squared test with a p-value very close to 0. The small p-value suggests that the observed difference in heater alarm rate among different seasons is not likely to occur by chance.

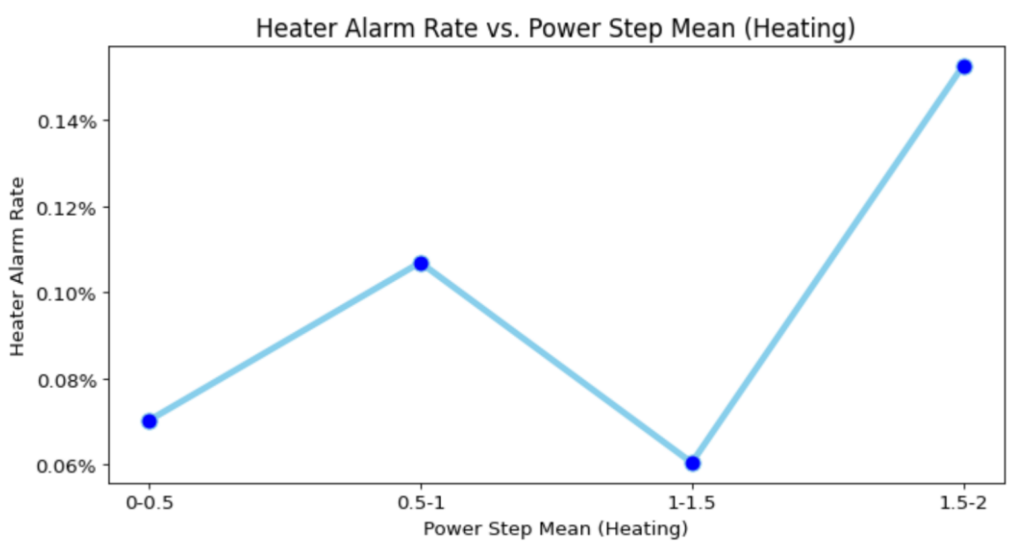
As ODTD increases, heater alarm rate increases as well. It seems that after total 10 hours of experiencing out of designed temperature, there is a big jump in the heater alarm rate. (See Figure 21)

Figure 21. *Heater Alarms Rate vs. ODTD*



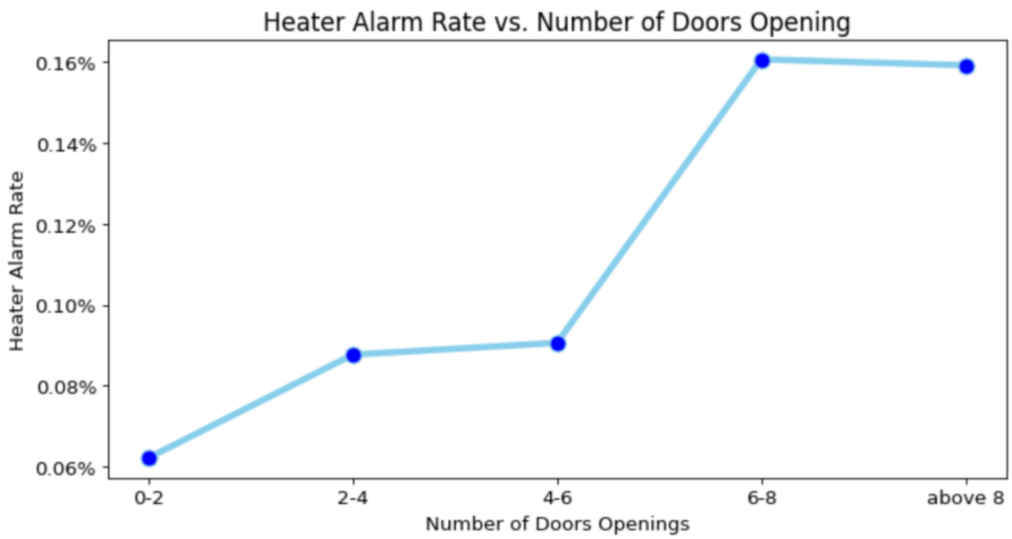
A higher power step mean suggests that the container is working hard to maintain its set temperature. According to our analysis result, the higher this power step mean is, the more likely the container is to trigger heater alarms. (See Figure 22)

Figure 22. *Heater Alarm Rate vs. Power Step Mean (Heating)*



Another factor affecting heater alarm rate is the number of doors opening. As the number of door openings increases, the heater alarm rate tends to climb, and when the number of door openings pass 4, there is big jump of the heater alarm rate. (See Figure 23)

Figure 23. *Heater Alarms Rate vs. Number of Doors Openings*

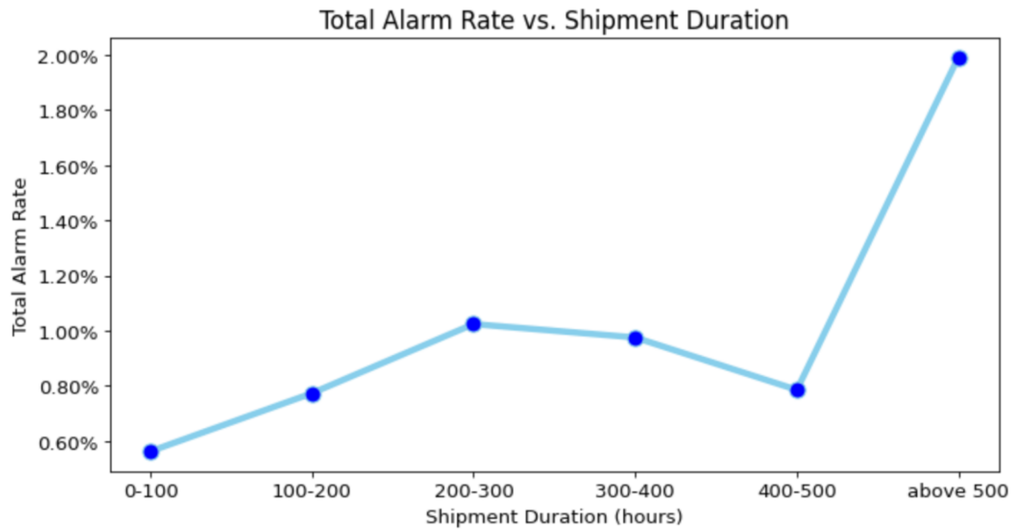


Total Failure Alarms

The following two factors show very similar trend against all three failure alarms, it would be more appropriate to analyze them against the combined count of all three failure alarms.

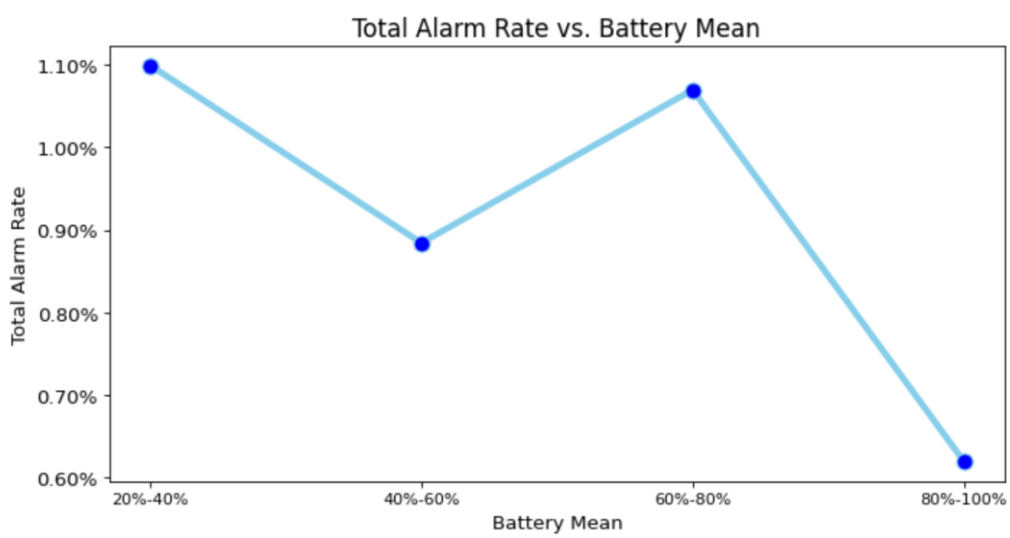
The first factor is shipment duration. The trend analysis demonstrates that the longer the shipment duration is, the higher the total alarm rate is, especially after the shipment duration reaches above 500 hours. (See Figure 24)

Figure 24. *Total Alarm Rate vs. Shipment Duration*



Battery mean is another critical factor influencing the total alarm rate. The trend analysis shows that the total alarm rate tends decreases as the battery mean increases. Maintaining a high battery level throughout the shipments can help to decrease total parts failure alarms rate. (See Figure 25)

Figure 25. *Total Alarm Rate vs. Battery Mean*



Compressor Performance Anomalies

After using Clayton copula and quantile regression to detect anomaly data points, we extract 552 under-performing data points and 1121 over-performing data points for further descriptive analysis. According to Table 4 below, we can see that the compressor alarm rate is 0.3% under normal instances, while it increases to 1.1% when the compressor is abnormally performing.

Table 4. *Compressor Performance Anomalies Factor Table*

	Compressor Alarm Ratio	Mean Battery Level	Mean Number of Door Openings	Charging Time Ratio	ODTD	TD Ratio
Normal	0.003	80%	3.5	0.47	0.045	0.13
Anomalies	0.011	71%	4.4	0.4	0.119	0.35

Battery level also plays an important role in the performance of compressor power step. Under normal instances, the container battery mean during shipment is 79.5%, while the battery mean is 71% when anomaly compressor performance occurs. Correspondingly, the charging duration ratio is lower when the anomaly compressor power step occurs. In summary, compressor performance anomalies tend to occur when container battery level is not properly maintained during a shipment.

Other than battery level, the number of door openings during the shipment is higher when the compressor performance is abnormal. Finally, when the compressor performance is abnormal, it tends to experience a longer out of designed temperature duration than normal instances and is also more likely to experience temperature deviation during the shipment. Long shipment duration and poorly maintained battery level can significantly increase the failure rate for all three parts, charger, compressor, and heater.

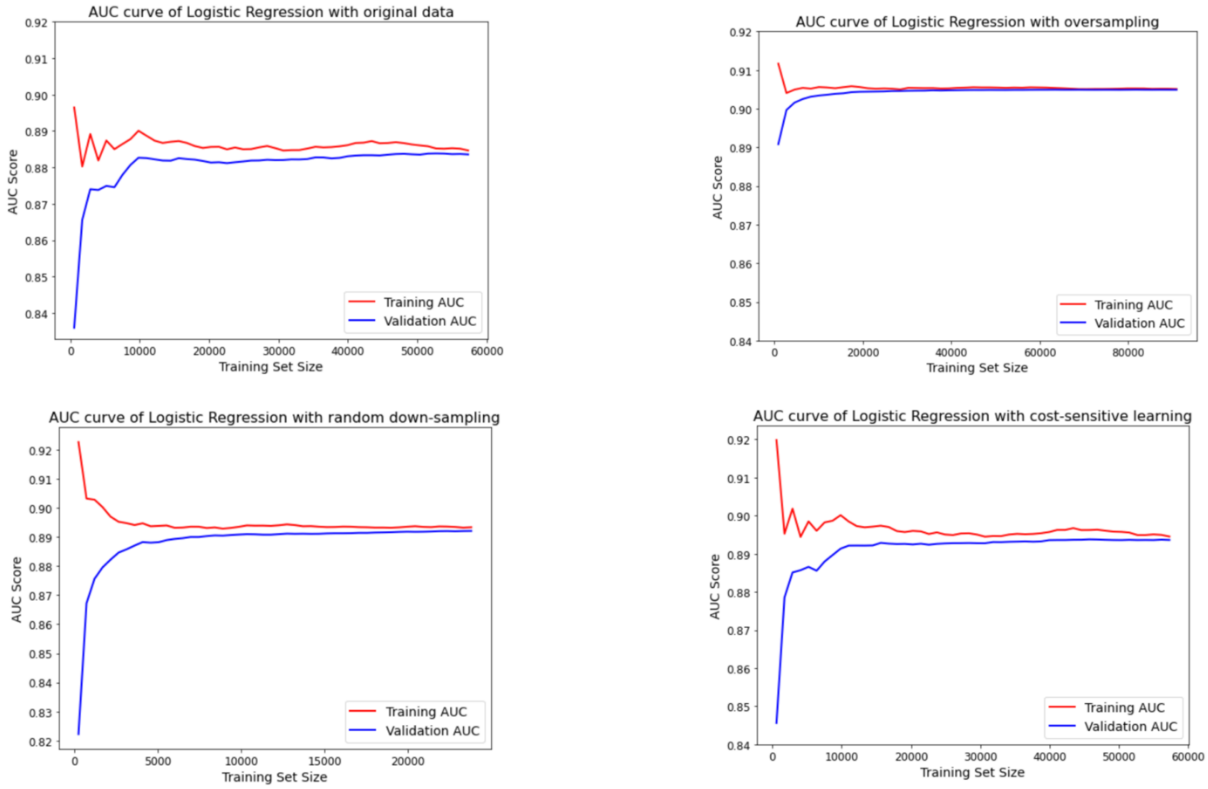
Predictive Model Findings

Logistic Regression, XGBoost, and Artificial Neural Network (ANN) with different imbalance data handling methods (adaptive sampling and cost-sensitive learning) were implemented.

- Original data: No adaptive sampling and cost-sensitive learning applied.
- Oversampling with SMOTE: Use SMOTE to oversample the minority class to reach the size of the majority class.
- Random down-sampling: Randomly down sample the majority class to reach the size of the minority class.
- Cost-sensitive learning: Change the weights of majority class and minority class in the cost function. In this case, majority class: minority class = 1:7.3.

There are a few additional considerations needs to be taken before the selection of a best model. First, model overfitting needs to be analyzed. Regardless of the validation AUC and recall scores, if a model shows a clear sign of overfitting through its training, that model needs to be discarded. Based on Figure 26, training and validation AUC scores converge nicely for logistic regressions with all imbalance data handling methods. There is no sign of overfitting for logistic regression.

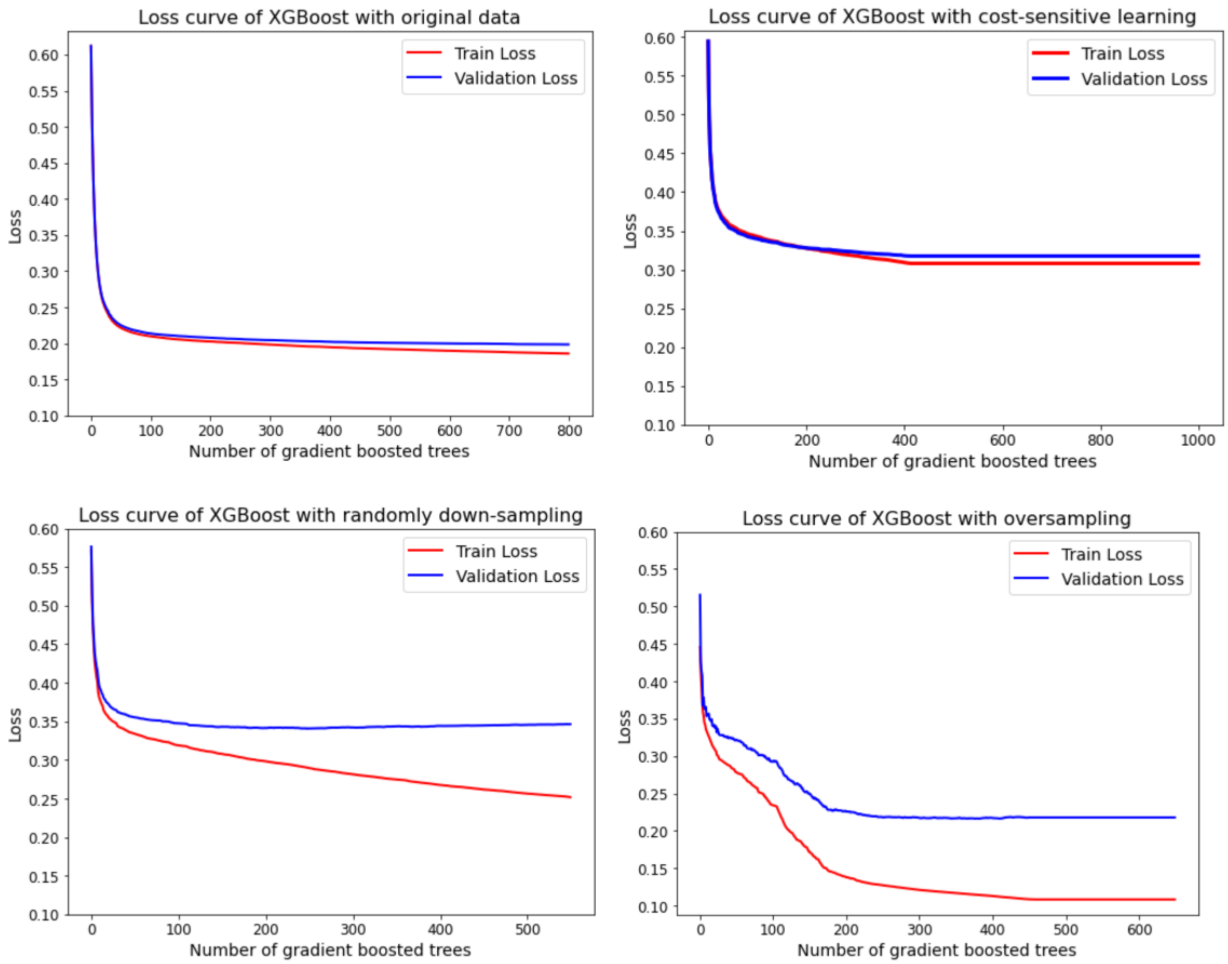
Figure 26. *Training and Validation AUC Score Progression of Logistic Regressions*

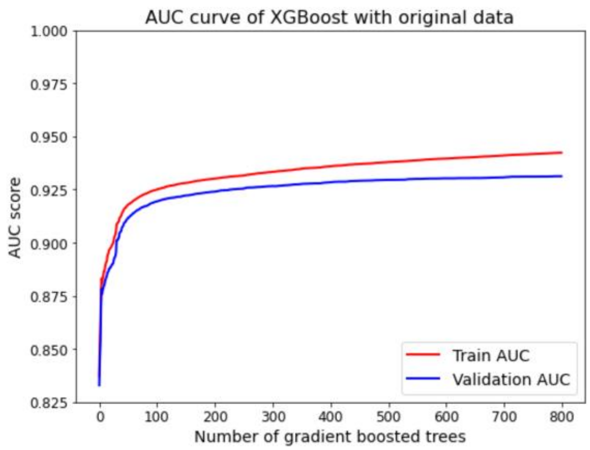


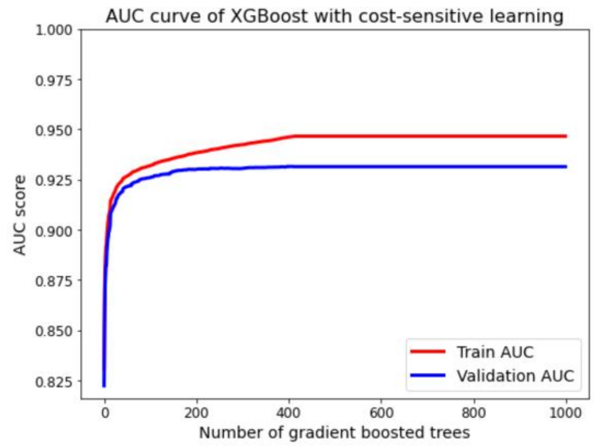
According to the loss and AUC score progression of XGBoost with different imbalanced data handling methods (Figure 27), over-sampling and down-sampling methods showed clear signs of overfitting, as there exists a huge gap between train and validation for both the loss curve and AUC scores. XGBoost with original data and cost sensitive learning showed much better results on train vs. validation loss and AUC. XGBoost with cost-sensitive learning is the best XGBoost model in terms of overfitting. However, there is still a little overfitting for the chosen XGBoost model.

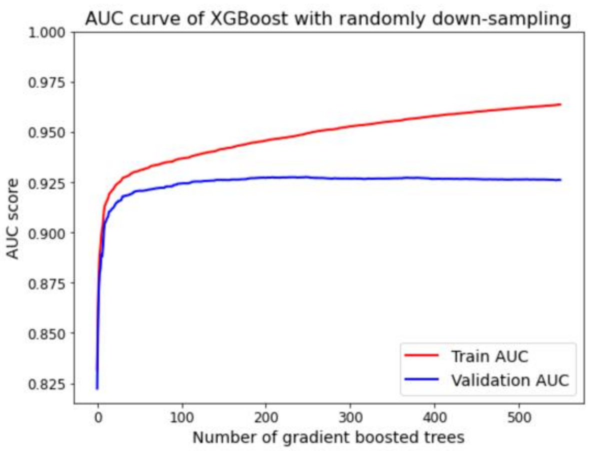
According to the loss and AUC score progression of ANN with different imbalanced data handling methods (Figure 28), all imbalanced data handling methods showed signs of overfitting except cost-sensitive learning. The validation loss for cost-sensitive learning is lower than the training loss, and both AUC scores converge very nicely. Thus, ANN with cost-sensitive learning is the best ANN model in terms of overfitting.

Figure 27. Training and Validation Loss and AUC Score Progression of XGBoost




AUC curve of XGBoost with original data
AUC score
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0
100
200
300
400
500
600
700
800
Number of gradient boosted trees
Train AUC
Validation AUC


AUC curve of XGBoost with cost-sensitive learning
AUC score
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0
200
400
600
800
1000
Number of gradient boosted trees
Train AUC
Validation AUC


AUC curve of XGBoost with randomly down-sampling
AUC score
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0
100
200
300
400
500
Number of gradient boosted trees
Train AUC
Validation AUC

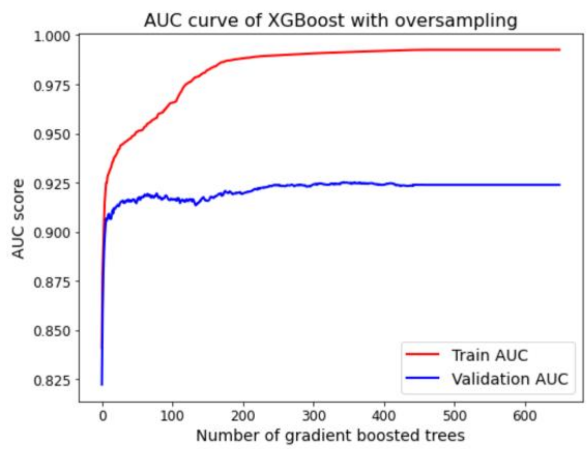
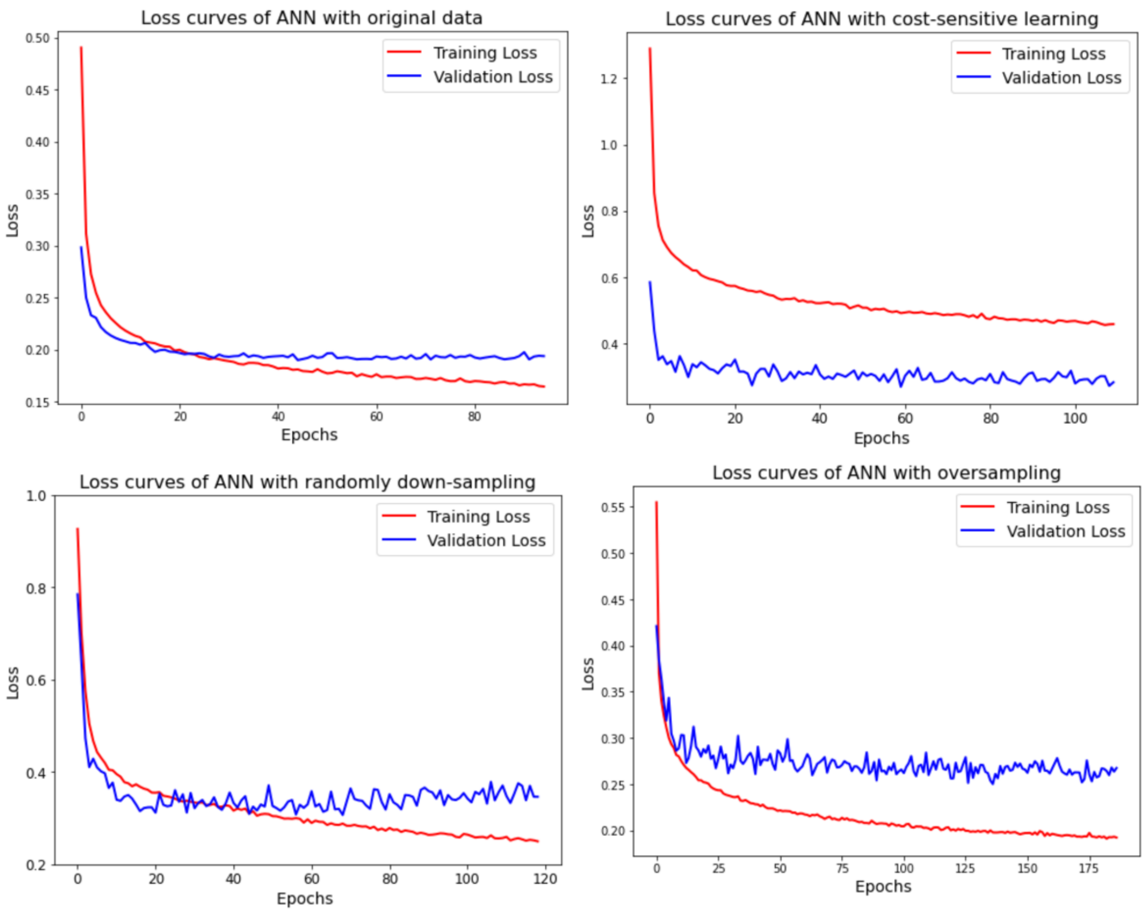

AUC curve of XGBoost with oversampling
AUC score
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0
100
200
300
400
500
600
Number of gradient boosted trees
Train AUC
Validation AUC

Figure 28. *Training and Validation Loss and AUC Score Progression of ANN*



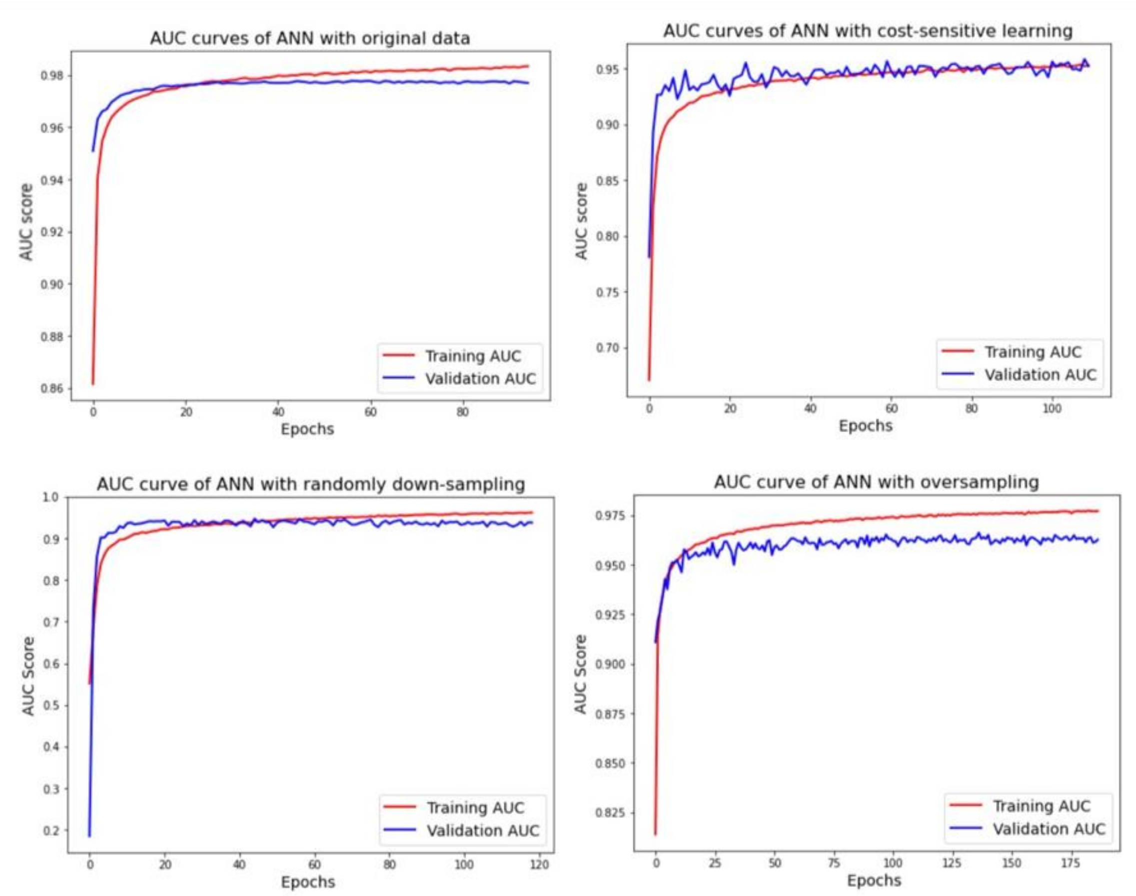


Table 5 and Table 6 record the AUC scores and recall scores for three different machine learning algorithms on original data and their combinations with different imbalanced data handling methods (adaptive sampling and cost-sensitive learning).

Table 5. *Validation AUC Scores*

AUC score	Logistic Regression	XGBoost	Artificial Neural Network
Original data	0.876	0.931	0.936
Oversampling with SMOTE	0.889	0.920	0.930
Random down-sampling	0.887	0.927	0.928
Cost sensitive learning	0.888	0.931	0.936

Table 6. *Validation Recall Scores*

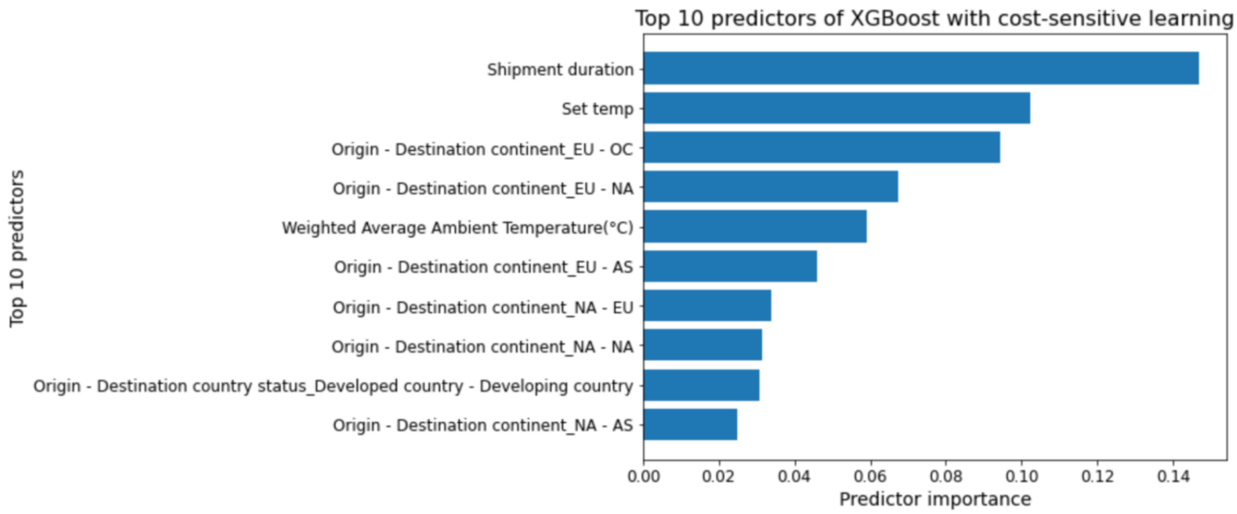
Recall score	Logistic Regression	XGBoost	Artificial Neural Network
Original data	0.298	0.486	0.484
Oversampling with SMOTE	0.821	0.590	0.825
Random down-sampling	0.820	0.878	0.891
Cost sensitive learning	0.821	0.878	0.891

Logistic regression and its combinations set up the baseline model performance. In terms of model overfitting, both XGBoost and ANN with cost-sensitive learning emerged as winners. Based on the AUC and recall scores, we can easily draw a conclusion that ANN results are better in validation performance.

Although it seems that models with original data performed well in AUC scores, it is actually an illusion created by the imbalanced data. In an imbalanced data set, the AUC score can be very high even if the minority class, temperature deviation, is not accurately predicted. Thus, we introduced recall score as another measurement. Although ANN with original data has a high validation AUC score of 0.936, it generated a much lower validation recall score of 0.484. Judging from both the validation AUC (0.936) and recall scores (0.891), ANN with cost-sensitive learning was chosen as the best model and further tested with the untouched test set. The result of a test set run on the chosen model produced an AUC score of 0.936 and a recall score of 0.891.

Although an ANN model was chosen as the best model due to its outstanding performance in overfitting, AUC and recall, we should still keep in mind that the interpretability of an ANN model is much less than that of an XGBoost model. The ranked feature importance of predictors used in a model can tell analysists which features have the most impact on the failure occurrences. XGBoost is able to produce such a list of feature importance (Figure 29), but ANN cannot. From the feature importance ranking, we can see that shipment duration and Delta are the two most important features that cause temperature deviations during a shipment.

Figure 29. *Feature Importance of XGBoost with cost-sensitive learning*



Summary and Conclusion

Temperature Deviation Risk Profile

The performance of E-containers is affected by the ambient temperature and their ability to maintain inside temperature. It is observed from our analysis that with a Delta (difference between average ambient temperature and set temperature during a shipment) over 10°C or below -10°C, containers are more likely to experience temperature deviation, and this kind of failure is especially common during summertime. Similarly, as long as the container is exposed to out-of-design ambient temperature (extreme temperature), the temperature deviation rate increases to 0.4. Since high delta causes containers to work harder, containers' workload also affects temperature deviations. As the average power output of a container increases to 2, the likelihood of temperature deviation will also increase. Moreover, handlers are advised to be extra cautious when the shipment duration increases to more than 10 days and when the shipment's origin country is a developing country.

The human handling behaviors is the other category of major temperature deviation causes. The more frequently handlers open the container door, the more likely containers will experience temperature deviation. This is because opening doors too often will expose containers more to the ambient temperature. It is suggested that the handlers avoid opening the door more than 5 times during a shipment. Another important factor affecting performance is container handler's ability to maintain a good battery level. Container handlers should maintain the battery level above 70% to significantly reduce temperature deviation rate. For E-containers, if the battery is not continuously and correctly charged, the temperature deviation rate will increase.

Airlines with high temperature deviation rates tend to exhibit poor container handling behaviors, e.g., not charging the container properly and maintaining a high battery level during shipments, opening container too frequently, as well as allowing containers exposed to extreme out of deign ambient temperatures.

Ambient temperatures in trade lanes with high temperature deviation rates tend to be much higher than container set temperatures. Consequentially, containers shipped through those trade lanes are usually exposed to extreme out of deign ambient temperatures for a large amount of time.

The comparison analysis between Self-Defined Temperature Deviations and Reported Temperature Deviations over non-human related factors shows very similar trends. This result validates the project's definition of temperature deviation.

Failure Alarm Risk Profile

Charger failures are twice as likely to occur when exposed to a supply voltage of 100V - 127V compared to a supply voltage of 220V-240V. Continuously charging the container and avoiding frequent alterations between charging and not charging are also essential to reducing charge failures. Once the charged duration reaches half of its shipment time, the handler should place extra precaution in the charging alarm.

Compressor and heater failures are more likely to occur when Delta (Difference between ambient temperature and set temperature) are large and containers are exposed to long extreme out of design temperature duration. Subsequentially, summer is a high season for both failure alarm occurrences. Minimizing the number of door openings to less than 5 times and avoiding constant large power output during shipments can also effectively reduce compressor and heater failures.

Compressor Performance Anomalies Profile

As the most utilized part of the container with the most occurrences of failure, compressor tend to have performance anomalies when container battery levels are not properly maintained, and higher number of door openings occurs during a shipment. Long exposure to extreme ambient temperature conditions can also cause compressor performance anomalies.

Predictive Model for Temperature Deviation

Both XGBoost and ANN with cost-sensitive learning performed well in terms of overfitting, AUC and recall. The best ANN model produced slightly better results in all three evaluation metrics. ANN with cost-sensitive learning is chosen to be the model for implementation. The chosen model can accurately predict close to 90% of temperature deviations. Base on the feature importance ranking, shipment duration and Delta have the highest impact on the occurrence of temperature deviations during a shipment.

Recommendations

Temperature Deviation

If containers experience high ambient temperature, certain actions should be taken to ensure the stability of containers inside temperature. Before shipping, it is suggested that shippers turn on the container no more than 2 hours before placing the product inside to avoid container experiencing long operation hour. The battery should be continuously charged, and the battery level should be regularly checked and maintained above 70% if possible. Putting containers into other temperature-controlled storages during transition is another effective way to keep containers away from the extreme out of design ambient temperatures. It is recommended that airlines provide extra temperature-controlled storage to place the container during transition. Number of door openings during a shipment should be minimized to reduce exposure to the ambient environment and workloads, under 5 times if possible.

Failure Alarms

During summer, Envirotainer should pay more attention to the E-container heaters and compressors, since the analysis result shows that both the number of heater and compressor alarms far exceed those of other seasons. In order to reduce charging alarm, shippers or airlines should also provide a good charging environment for the container throughout the trip. Also, when the supply voltage of either origin or destination country is between 100V - 127V, it is recommended to use voltage transformers. Envirotainer should consider a charger design modification to better adapt supply voltages ranging between 100V and 127V.

Compressor Performance Anomalies

According to the compressor power step anomaly detection analysis, it is essential to maintain the high battery level to avoid compressor power steps from under-performing or over-performing. Therefore, shippers, airlines, and forwarders should not let the container’s battery less than 30% duration takes up more than 9% of the total shipment duration and should maintain an average shipment battery level above 70%.

Failure Prediction Model

Before a shipment starts, it is recommended to run the chosen predictive model utilizing all the available shipment information to produce a likelihood of temperature deviation for that shipment. If the likelihood is higher than a critical percentage, extra caution and preventative measures mentioned above should be placed to avoid temperature deviations. Close attention needs to be placed on a shipment, if the projected shipment duration is long and the difference between forecast ambient temperature and set temperature is large.

References

Aggarwal, C. C. (2017). *Outlier Analysis*. Springer, Cham.

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1). https://doi.org/10.1186/1471-2105-14-106

Cece, S. (2020, April 6). *Pharma needs a smarter cold chain*. Pharma Manufacturing. https://www.pharmamanufacturing.com/articles/2020/pharma-needs-a-smarter-cold-chain/.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Comes, T., Bergtora Sandvik, K., & Van de Walle, B. (2018). Cold chains, interrupted: The use of technology and information for decisions that keep humanitarian vaccines cool. *Journal of Humanitarian Logistics and Supply Chain Management*, *8*(1), 49–69. https://doi.org/10.1108/jhlscm-03-2017-0006

Grand View Research. (2021). (rep.). *Pharmaceutical Logistics Market Size, Share & Trends Analysis Report By Type (Cold Chain Logistics, Non-cold Chain Logistics), By Component (Storage, Transportation, Monitoring Components), By Region, And Segment Forecasts, 2021 - 2028*. Retrieved from https://www.grandviewresearch.com/industry-analysis/pharmaceutical-logistics-market

Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.35.1693&rep=rep1&type=pd f

PRESS RELEASE PR Newswire. (2018, January 31). *Envirotainer Receives Award as Best Active Temperature-Controlled Packaging Solution Provider*. Markets Insider.

https://markets.businessinsider.com/news/stocks/envirotainer-receives-award-as-best-active-temperature-controlled-packaging-solution-provider-1014696018.

STAT Times. (2020, November 20). *Envirotainer successfully helps transport Covid-19 vaccine Sputnik V to Abu Dhabi*. THE STAT TRADE TIMES. https://www.stattimes.com/news/envirotainer-successfully-helps-transport-covid19-vaccine-sputnik-v-to-abu-dhabi-logistics/.

VRR. (2020, August 25). *The Role of Cold Chain Logistics in the Pharmaceutical Industry*. https://vrr.aero/this-is-vrr/news/2020/the-role-of-cold-chain-logistics-in-the-pharmaceutical-industry/.

Yoon, J., Jordon, J. & Schaar, M.. (2018). GAIN: Missing Data Imputation using Generative Adversarial Nets. Proceedings of the 35th International Conference on Machine Learning, in Proceedings of Machine Learning Research 80:5689-5698 Available from http://proceedings.mlr.press/v80/yoon18a.html.