Chiller parts expected fault date prediction and visualization

Yuxin Ma
Bangguo Wei

August 27, 2022

Contents

1   Business background

The company finds it's costly to arrange a worker to repair a chiller without knowing which parts are failed inside. The worker will have an issue identifying what parts to bring when scheduling their repair job and is time-consuming to fetch the broken parts twice. Only (53) percent of all repairs that required component parts were wrapped up with a single visit, and It costs about ($6.49) to return every single unnecessary part to a warehouse. The algorithms in this project are aimed to identify what parts failed during a period and to offer reliable guidance to repair workers. The project is set up to mitigate the inefficiency issue just raised above, thus improving the workers' productivity and improving the company's profit in the long run.

2   Machine Learning model

2.1   Framework

2.1.1   Data

The original data set is confidential and is provided by the business team. The records contain 2 sample chillers, one chiller includes 43 parts, and the other includes 34 parts. For each part, we could obtain part description, identical part code, quantity, expected functioning hours and expected life cycle in years.

2.1.2   Preprocess

Before fitting into the models, we first expanded the parts that have more than one quantity. Then for more direct visualizations, a randomized start functioning date is simulated.

2.1.3   Plan

Predicted parts' life cycles are added up to each part's start functioning date.

2.2   Models

We use the lifelines library in the Python library to analyze the product life distribution and predict the likelihood of an incoming replacement. So it's helpful for the company to prepare enough inventory and maintenance teams in advance.

2.2.1   Training data labels

For this supervised training, we need labels for training data. We set parts' life to label the original features data. We will give higher life scores when the outcome of the prediction is better.

2.2.2   Kaplan-Meier survival analysis

Kaplan-Meier curve is a non-parametric statistic used to estimate the survival function from lifetime data, and we use the Kaplan-Meier curve to display how survival rates of the parts have changed in the figure 1 below. The horizontal axis in the figure 1 represents time, the vertical axis on the left represents survival probability, and the vertical axis on the right represents cumulative density. The figure shows that the median survival time of parts is 1269 days and the median survival confidence interval is between 865 and 1486 days.
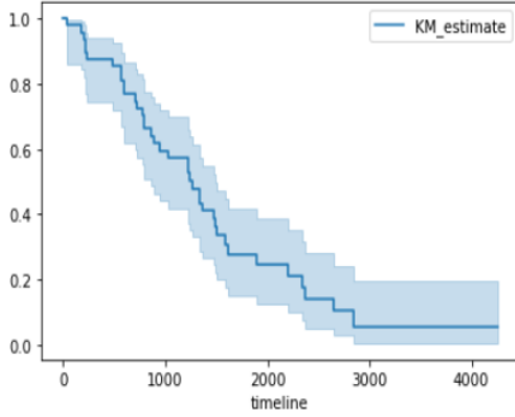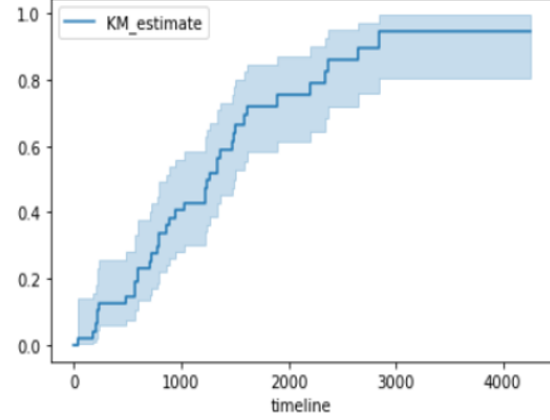
(a) Survival Function   (b) Cumulative Density

Figure 1: Kaplan-Meier curve for Parts life

2.2.3 Wei-bull distribution

In probability theory and statistics, the Weibull distribution is a continuous probability distribution that can fit an extensive range of distribution shapes. So we also use the Weibull distribution to fit the survival time probability of the part in Figure 2 below. The horizontal axis in figure 2 represents time, the vertical axis represents survival probability. The result in figure 2 shows the median survival time of parts is 1216 days and the Weibull mean survival time is 1486 days. The result is very similar to the Kaplan-Meier survival analysis.
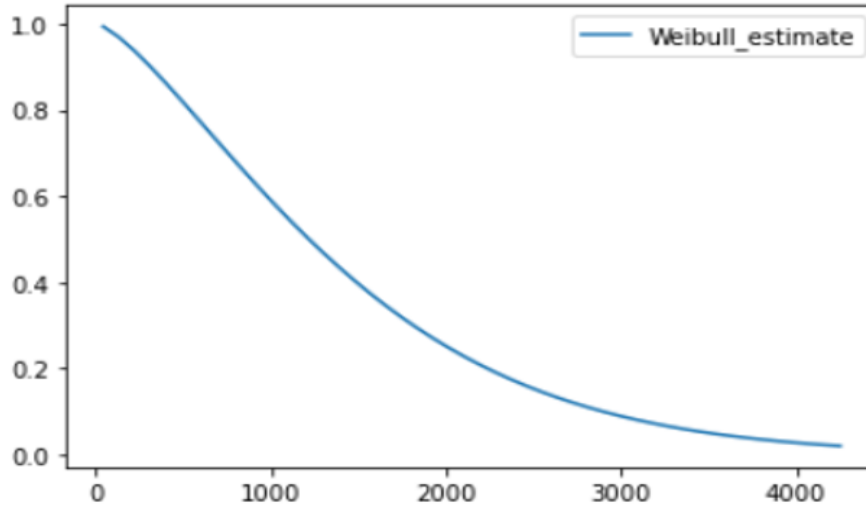

Figure 2: Weibull distribution estimate

2.2.4 Model extraction

When predicting part life, the output of the Weibull model is that each part increases with time, and the probability of part survival changes, so by setting a probability threshold, we can predict whether a part is normal or at risk of damage. On the other hand, we can adjust the probability threshold to

3

control the range of parts that are detected as risky. To evaluate our model, we use the AIC (Akaike information criterion) in the Lifelines library to judge the performance of the model, and perform statistical analysis (as shown in the figure below).

model	lifelines.WeibullAFTFitter	
duration col	'Time to Replace'	
event col	'Ended'	
number of observations	48	
number of events observed	38	
log-likelihood	-314.869	
time fit was run	2022-07-04 00:18:05 UTC	

		coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
lambda_	Intercept	7.368	1584.779	0.118	7.138	7.599	1258.532	1995.599	0.000	62.653	<0.0005	inf
rho_	Intercept	0.326	1.386	0.130	0.072	0.581	1.075	1.787	0.000	2.516	0.012	6.399

Concordance	0.500
AIC	633.738
log-likelihood ratio test	-0.000 on 0 df
-log2(p) of ll-ratio test	nan

Figure 3: Weibull model fitting results

2.3 Visualisation with Dash

2.3.1 Introducing Dash

We visualized the expected fault time for each part using the Dash package in python. The interactive dashboard by Dash package allows users to check their targeted information by simply hovering the mouse and filtering.

2.3.2 Functions

The expected parts' life cycles are obtained from the Wei-bull distribution described above. After adding on the start functioning date of each part, we could visualize the predicted fault date as in Figure 4 below. The x-axis shows when a part is about to fail while the y-axis indicates the identical part code. We could click on the dots in the graph to see each part's details, including part code, which chiller it belongs to, expected fail date, and description.
The table below the plot contains detailed information associated with each part. It clarifies the details, and thus offers a more precise and translucent display to the users by scrolling down the cursor.

The text box below the plot allows users to input a certain month with the format 'YYYY/MM'. Click on the 'Check' button, it can return which parts are expected to fail at the indicated point of time. The following figures show two possible scenarios when users input months correctly.

3 Business impact

The business impact of the model is estimated as follows.
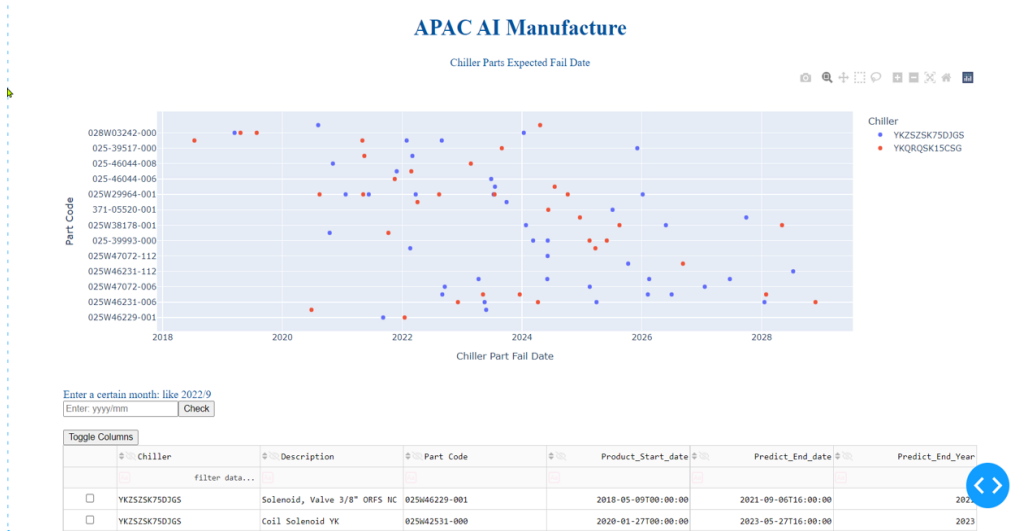**Savings**: When a part reaches a critical life, we can try to monitor it through the dynamic interactive

Figure 4: Chiller parts fault date overview

(a) Parts about to fail
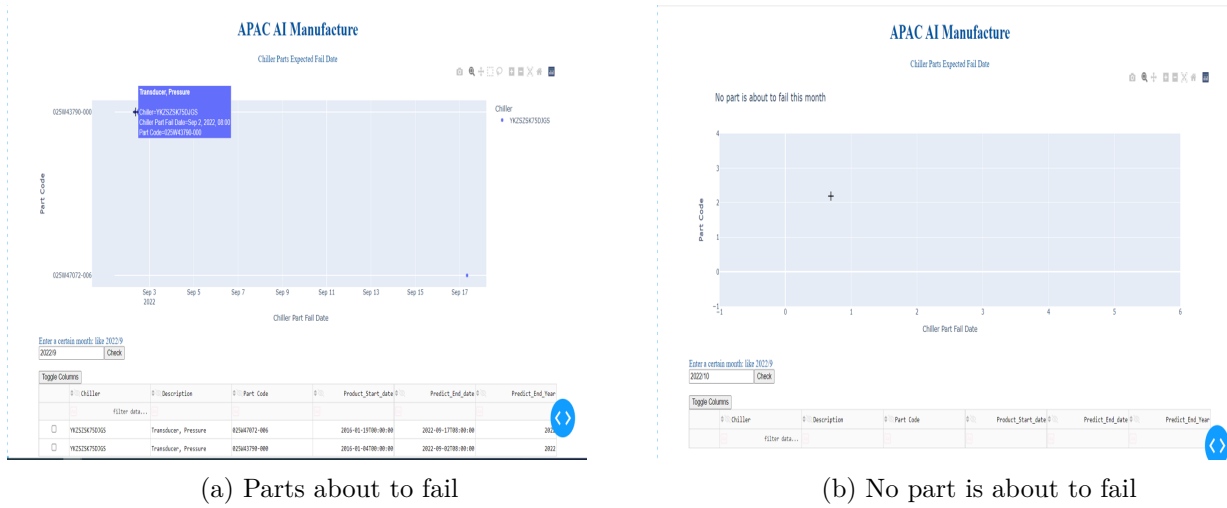

(b) No part is about to fail

Figure 5: Two possible scenarios with user input

web page, check the life of the part in time, and provide effective suggestions for the exchange of parts (Note: the results are adopted by the commercial sector).