The Need for Standardized Metrics and KPI's for AI Performance

Rory S. Smith

University of Northampton, Northampton, UK

Keywords: Artificial Intelligence, Metrics, Maintenance.

Abstract. The use of Artificial Intelligence (AI) in the lift industry is becoming commonplace. It, along with remote monitoring, is being applied to lift and escalator maintenance.

Some governments who require monthly or twice monthly maintenance now only require quarterly maintenance if remote monitoring and AI are utilized.

The benefits of AI-augmented maintenance such as increased up time, improved first time fix rate and fewer running on arrival (ROA) calls are being touted by governments and lift companies alike. However, no standardized set of metrics exists for these benefits.

A set of metrics for lift and escalator maintenance is proposed and discussed.

1 INTRODUCTION

In this paper, existing service metrics for lifts and escalators will be reviewed. The meanings of these metrics and how these metrics are used to evaluate service performance when using Artificial Intelligence (AI), as well as how these metrics can be misinterpreted or even manipulated, will be reviewed.

2 METRICS

2.1 False Positives, False Negatives

Machine Learning based on Pattern Recognition is one form of AI that is being used for lifts [1]. One of the goals of Machine Learning (ML) is to be able to predict that a lift will fail in the near future and send a technician to the site to fix the lift before it fails.

Vibration patterns of many lift components can be captured by accelerometers. The vibration of a new component will be different from an older or damaged component. The remaining life of these components can be monitored and these components can be replaced before they fail.

When a prediction is made that a lift WILL fail there are two possible outcomes as follows:

1. True Positive (TP). It was predicted that the lift would fail, and it fails.
2. False Positive (FP). It was predicted that the lift would fail, and it does NOT fail.

When it is predicted that a lift will NOT fail there are also two outcomes:

1. True Negative (TN). It was predicted that the lift would not fail, and it did not fail.

2. False Negative (FN). The lift was predicted not to fail, and it failed.

The following example is used to explain the terms False Positives and False Negatives as well as True Positives and True Negatives [2].

Example 1. A Building Complex with 100 lifts equipped with IOT remote monitoring devices are connected to a cloud-based predictive analytics system. This system has predicted that 10 of the 100 lifts will fail in the next 2 weeks. This also means that the other 90 lifts will not fail.

A technician is sent to each of the 10 units that are predicted to fail and finds 8 lifts that have a condition that if not repaired will definitely cause the lift to fail in 2 weeks. He fixes the 8 lifts.

During the next two weeks, the 2 lifts that were predicted to fail do NOT fail. However, 3 other lifts do fail.

Of the 10 predictions, there were 8 True Positive predictions and 2 False Positives.

Additionally, there were 3 False Negative predictions.

One method of visualizing these results is a Confusion Matrix [3]. Figure 1 is a confusion matrix of these results.

		Predicted	
		Failure 10	No Failure 90
Actual	Failure 11	TP 8	FN 3
	No Failure 89	FP 2	TN 87

Figure 1 Confusion Matrix

These results will be viewed differently by the building complex manager than by the Lift Company's branch manager.

Building Complex manager's perspective:

> Prior to adding the IOT equipment the complex would have experienced 11 breakdowns. The IOT system prevented 8 breakdowns. The complex manager is satisfied with the improvement.

Branch manager's perspective:

> With or without IOT, 11 lifts would need to be repaired. With IOT, 2 additional service visits to the lifts that did not fail were required. The branch manager is unhappy because the two visits to lifts that did not fail will reduce the branch profitability and his bonus.

False Positives are an important metric. The lower the occurrence rate of False Positives, the happier both the Building Manager and the Branch Manager will be.

Prior to installing IOT equipment, all breakdowns were False Negatives. Fixing a lift before it fails should come at a lower cost, particularly if it can be repaired during normal working hours.

Eliminating False Negatives should improve both customer satisfaction and operational efficiency. False Negatives can also be labelled *Unpredicted Failures*.

Unpredicted failures, particularly the reduction of unpredicted failures, is a significant metric.

2.2 Precision and Recall

Statisticians use the metrics of Precision and Recall in evaluating predictions [4].

Precision represents the accuracy of the predictions that were made. It is defined by the following formula:

$$Precision = \frac{Tp}{(Tp+Fp)} \tag{1}$$

Where: Tp represents True Positives

Fp represents False Positives

In Example 1, ten predictions were made. There were 8 True Positives and 2 False Positives. The Precision of these predictions was 0.8.

Recall represents the number of True Positives that were predicted compared to the number of True Positives that could have been predicted. Recall is defined by the following formula:

$$Recall = \frac{Tp}{(Tp+Fn)} \tag{2}$$

Where: Tp represents True Positives

Fn represents False Negative

Ideal Values of both Precision and Recall are 1.0.

In Example 1, Eight True Positives were predicted. However, there were a total of Eleven Positives. Since there were 3 Failures that were not predicted, there were 3 False Negatives. The Recall value of these predictions was 0.73.

Recall and Precision are not metrics of the Predictive Maintenance system but rather are metrics of the Machine Learning system that is a part of the Predictive Maintenance system. If the ML system has Recall and Precision values of 1.0 but they are ignored and no actions are taken to fix the lift before it fails, then the Predictive Maintenance system has no value to either the building or the lift maintenance provider.

It should also be noted that if the ML algorithm is conservative, then there can be high Precision levels because there will be no false positives. However, Recall values will be low because false negatives will be high.

2.3 Lift Failures

What is the definition of a Lift Failure or the more commonly used term, Breakdown?

One definition proposed is, "The lift is unable to move the users up and down in a building" [5]. This seems to be a simple definition. However, consider the three following examples:

1. There are some failures that appear for a period of time and then the lift returns to service without any human intervention. These failures often go unreported.

2. Buildings report a lift is out of service, but when the technician arrives, the lift is functioning properly. This type of failure is often reported as Running on Arrival or ROA.

3. A Machine Learning system connected to a lift determines the lift is out of service and initiates a remote reset of the controller. After the reset, the lift functions normally. The building did not detect that the lift was out of service.

Should any of these three examples be considered failures? What is good for a brochure might not be so good for the building operator. All failures should be reported so the AI system can be properly evaluated.

2.4 MTBF

MTBF, Mean Time Between Failures is a standard metric used for years in industrial applications [6]. Since one of the benefits of AI is the ability to fix things before they fail, AI should result in longer MTBF values. MTBF can be calculated as follows:

$$MTBF = \frac{\sum(start\ of\ downtime - start\ of\ uptime)}{number\ of\ failures} \tag{3}$$

2.5 MTTR

MTTR, Mean Time to Repair is a metric that simply calculates the average time required to return a lift to service that has failed. AI should shorten the time to repair because simple repairs can be performed before the failure becomes more serious and additionally reduces diagnostic times.

$$MTTR = \frac{\sum(start\ of\ uptime - start\ of\ downtime)}{number\ of\ failures} \tag{4}$$

2.6 Failures per Unit

Failures per Unit are usually expressed as failures per unit for a time period such as a year. The average number of failures or breakdowns per year is difficult to determine from published literature. One source stated that breakdowns per unit per year of well-maintained lifts varies from 0.5 to 2 per year [7]. Another source states that the average lift has 4 breakdowns per year [8]. A major manufacturer claims to have less than 1 breakdown per unit per year on the lifts that they maintain [9].

The ability of AI to predict impending failures before they occur and permit lifts to be fixed before they fail should reduce the failures per unit per year.

2.7 First Time Fix Rate

First Time Fix Rate (FTFR) is a metric used to determine the efficacy of an unscheduled service visit [10]. How is a First Time Fix defined? One definition is that the fault does not reoccur for 30 days [5]. The following are two examples that could make the reported FTFR appear better than the true FTFR.

Example 1.

Sequence of Events:

Technician A is dispatched to the site of a breakdown. The lift is not running, and he can find no apparent reason for it being out of service. He cycles the isolator switch, and the lift starts to run. On his service ticket he reports "Found lift out of service, adjusted door locks on floors 1, 2, and 3. Car returned to service".

Two weeks later the lift has a second breakdown. Technician B is dispatched to the same site, and he also cannot find the cause of the problem. He cycles the isolator switch, and the lift starts to run. On his service ticket he reports "Found lift out of service, adjusted Gate Switch and cleaned car door tracks. Car returned to service".

Three weeks after the first breakdown, Technician C was dispatched to the site to upgrade the controller software. The lift has not been reported out of service and Technician C finds the lift running properly. The technician installs the new software, and the lift is still functioning properly after 90 days.

Impact on FTFR:

Although the two breakdowns occurred within 30 days, because the problem resolutions were different, both breakdowns were reported as being Fixed the First Time. Because a software upgrade prevented further breakdowns, the root cause of the breakdowns was a software bug. The two breakdowns were not Fixed the First Time. However, they were reported as being Fixed the FirstTime.

Example 2

Sequence of Events:

A hydraulic lift is serviced quarterly. When the piston leaves the cylinder as the lift travels in the up direction, a film of oil is deposited on the piston. As the piston returns into the cylinder, the film of oil is wiped off the piston and into a collector ring. A small hose is connected to the collector ring which transports the oil into a 19-litre container.

A technician is dispatched to the hydraulic lift site because the lift is reported out of service. The technician finds the lift shut down due to a lack of oil. He empties the oil from the overflowing 19-litre oil container into the oil tank and adds 10 additional litres of oil to the tank. The lift is returned to service.

Six weeks later the lift is shut down again, and again the problem is a lack of oil. This time a repair crew is sent to replace packing gland seals and add 20 litres of oil. With new seals, the lift will deposit a smaller film of oil on the piston. The lift can now run without shutting down due to low oil between the quarterly services.

Impact on FTFR:

Because the shutdowns occurred more than 30 days apart, the first breakdown was reported as being Fixed the First Time. However, the lift was not fixed until the packing was replaced after the second low-oil event.

In both of these two examples, the lift was reported as Fixed the First Time, making the FTFR metric look good. However, the breakdowns per unit per year metric will look worse than if the problems were fixed the first time.

2.8 Running on Arrival Incident Rate

When a service technician is dispatched to repair an out-of-service lift, it is very common to find the lift Running on Arrival. There is the assumption that there was never anything wrong with the lift and that the building management company failed to verify that the lift was truly out of service.

One lift company will invoice the customer for ROA calls unless they have an IOT-based remote monitoring system [11].

Those with experience in service operations have found that ROA calls usually involve a real problem that is intermittent, or condition based.

For example, a door lock may be misadjusted. The lift will function properly most of the time. However, if passengers are distributed in the lift car in a particular way, the car may tilt slightly and cause the door to not properly lock. In this case, the car will not run until the passengers exit and the car balance changes. With the change in car balance, the car then runs. This creates an ROA event.

Are ROA calls considered a breakdown? In many cases, they are not counted and the customer is invoiced for a nuisance call.

The number of ROA calls could be a performance metric. A large number of ROA incidents indicates there are unresolved intermittent problems.

3 CONCLUSIONS

There are metrics available that can effectively be used to evaluate the efficacy of the application of AI in lifts and escalators. However, the definitions of what is a breakdown, what is a First Time Fix, or the significance of an ROA call can have an impact on the meaningfulness of these metrics.

A set of standardized metrics would be helpful for consultants and building management when considering an investment in AI and for evaluating the efficacy of an installed AI system. Additionally, standardised metrics would help the lift companies better understand the efficacy of their AI systems and how to improve those systems.

4 LITERATURE REFERENCES:

[1] S. Kaczmarczyk, R. Smith, Mateusz Gizicki *Vibration Signature and the Application of Intelligent Pattern Recognition in Detection and Classification of Damage in Automatic Power Operated Lift Doors* Available from: https://liftescalatorlibrary.org/paper_indexing/papers/00000500.pdf Last Accessed: 27 July 2023.

[2] *False Positives and False Negatives* Available from: https://en.wikipedia.org/wiki/False_positives_and_false_negatives Last accessed: 6 June 2023.

[3] *Confusion Matrix* Available from: https://en.wikipedia.org/wiki/Confusion_matrix Last accessed 27 July 2023.

[4] *Precision and Recall* Available from: https://en.wikipedia.org/wiki/Precision_and_recall Last Accessed: 27 July 2023.

[5] Yao, C. *The Use of IoT Technologies in Condition-Based Maintenance of Lifts – Setting Requirements in Singapore* University of Northampton, Northampton, UK (2023)

[6] *Mean Time Between Failures* Available from: https://en.wikipedia.org/wiki/Mean_time_between_failures Last accessed:8 June 2023

[7] *How reliable are Elevators and Escalators* Available from: https://elevatingstudio.com/how-reliable-are-elevators-and-escalators/ Accessed: 13 June 2023.

[8] *What Causes 272 million hours of down-time each year* Available from: https://www.twinfm.com/article/what-causes-272-million-hours-of-downtime-each-

year#:~:text=New%20data%20from%20smart%20elevator%20technology%20start-up%20Uptime,nearly%20272%20million%20hours%20of%20downtime%20each%20year. Last accessed: 5 June 2023.

[9] *Mitsubishi Service* Available from: https://www.mitsubishielevator.com/uploads/files/pdf/ME_EED_Service_Brochure_May2023.pdf Last accessed: 9 June 2023.

[10] *First Time Fix Rate* Available from: https://limblecmms.com/blog/first-time-fix-rate/ Last accessed: 8 June 2023.

[11] *Schindler Ahead Fact Sheet* Available from: https://www.schindler.com/content/dam/website/us/docs/ahead/schindler-ahead-fact-sheet.pdf/_jcr_content/renditions/original./schindler-ahead-fact-sheet.pdf

BIOGRAPHICAL DETAILS

Rory Smith is Visiting Professor in Engineering/Lift Engineering at the University of Northampton and a Consultant at Peters Research Ltd. He has over 54 years of lift industry experience during which he held positions in research and development, manufacturing, installation, service, modernization, and sales. His areas of special interest are Machine Learning, Traffic Analysis, dispatching algorithms, and ride quality. Numerous patents have been awarded for his work.