DATA-DRIVEN ROOT CAUSE ANALYSIS IN HEAVY EQUIPMENT WITH ADVANCED ANALYTICS AND AI

Introduction

In today's data-driven world, understanding and resolving problems efficiently is crucial for industries like heavy equipment, trucking, and rail – where downtime, safety, and reliability directly impact productivity and profitability. Root cause analysis (RCA) is a systematic approach to identifying the underlying reasons behind issues and anomalies, helping organizations improve quality, optimize operations, and mitigate risks. While many companies already perform some form of RCA, leveraging advanced tools and methodologies can enhance the process.

This white paper explores key stages of RCA, discusses various analytical techniques, and highlights how data analytics and artificial intelligence (AI) platforms like Altair® RapidMiner® can drive more effective data-driven investigations.

What is RCA?

RCA is probably something your organization already does. At its core, RCA is a process of discovering the sources of problems and anomalies and solving them. RCA is a critical process in heavy equipment, truck, and rail operations:

- **Manufacturing**: Discovering the cause of faulty parts from machine data.
- **Supply Chain**: Analyzing bottlenecks and vulnerabilities.
- **Equipment and Infrastructure**: Diagnosing why a truck, rail car, or piece of heavy machinery failed, malfunctioned or experienced reduced performance.
- **Operational Performance**: Uncovering driver/operator errors, scheduling issues, or procedural breakdowns that affect efficiency or safety.
- **Fleet and Operator Safety**: Investigating accidents or mechanical issues to improve safety practices and reduce future risk.
- **Regulatory Compliance**: Ensuring adherence to transportation, environmental, and safety regulations by identifying and correcting non-conformities.
- **Maintenance and Sustainment**: Pinpointing maintenance issues, readiness shortfalls, or unexpected lifecycle costs to ensure long-term effectiveness.

RCA is crucial for organizations of all kinds, in every industry. It helps improve the quality of products, enhancing customer satisfaction and retention, and can even avoid health disasters, production delays, and product failure. However, it's important to realize that the RCA process isn't one-size-fits-all, meaning it can be optimized. With the right tools, organizations can make their investigations more rigorous, more systematic, and more automated – and they can apply this new process across business architectures regardless of data complexity.

When approaching RCA, there are several common frameworks that can be adopted. The “5 Whys” technique involves repeatedly asking “why” to trace a problem back to its root cause. Failure mode and effects analysis (FMEA) systematically assesses potential failure points and their impacts, while Pareto analysis uses the 80/20 rule to prioritize issues’ most significant causes. No matter what method or approach is chosen, analysis tools must be flexible, powerful, accessible, and scalable.

Step 1: Access

Everything starts with data. RCA relies on performing analysis on accurate, complete data; if we only have half the data, then we only have half the answer. Unfortunately, data quality is often a significant obstacle. This can be something as simple as data captured in an inconvenient format, such as unstructured documents like PDFs. The first step to effective RCA is to quickly convert data into consistent, easy-to-use formats like relational tables and spreadsheets. However, this is only the tip of the data quality iceberg.

When dealing with problems that span multiple departments or business units, organizations want to be able to link different sources of information to build a single pane of glass. For example, imagine you’re responsible for investigating repeated equipment failures across a national rail operator’s fleet. You’ve been provided with a list of related incidents which may represent some sort of systematic issue and are tasked with determining the root cause. The source could be anything: a problem during final assembly, issues at a specific maintenance depot, failures isolated to a particular region or specific manufacturing run, or even with the initial design.

Ultimately, collecting and connecting data from across the product life cycle is critical for these kinds of large-scale investigations. A data fabric enables the creation of a lightweight, universal overlay across all your inputs. Rather than making new copies, the data is left at source, and you have complete data lineage and traceability to support your investigation.

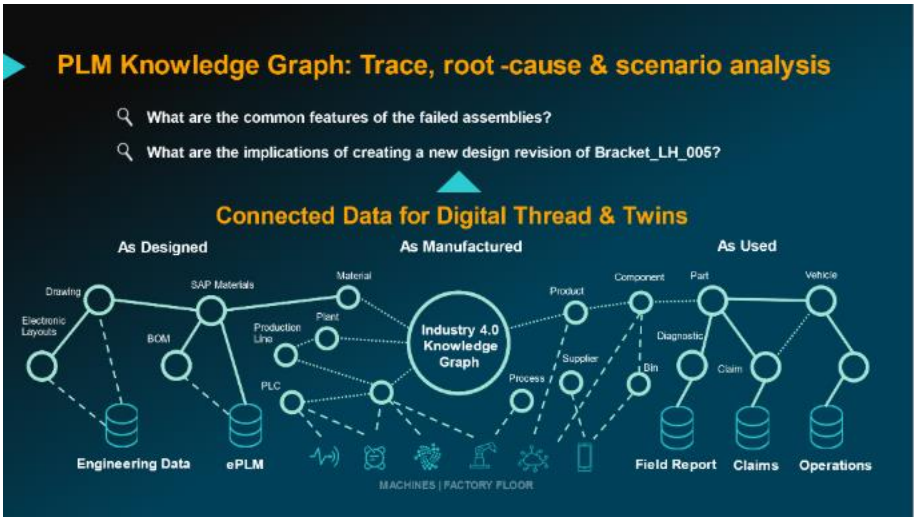

Figure 1 – One common use case is to combine a product life cycle management (PLM) tool with additional data to combine data from the entire life cycle of a product.

Regardless of type and format, your data is connected into a single network, and any new data sources can be connected in the future without any rework. The data can be harmonized to a consistent data model and blended to present different views to facilitate the stage of the RCA process.

Step 2: Analysis

Analysis is where we begin to profile and extract insights from data. Historically, engineers and analysts may have used spreadsheets as an easy way to perform analysis, but this method presents a few problems. Primarily, these spreadsheets can be tricky to manage and use, with poor scalability as the size of datasets grows. They can also rely on hard-to-understand, error-prone formulas and hide advanced features behind coding.

The alternative to this is using coding languages like Python, but not everyone has the skills to do this. Moreover, as coding comes into play, complexity increases, and it can be hard to manage for more than individual users. Fortunately, democratized data science that solves these issues isn't merely a dream – it's available right now.

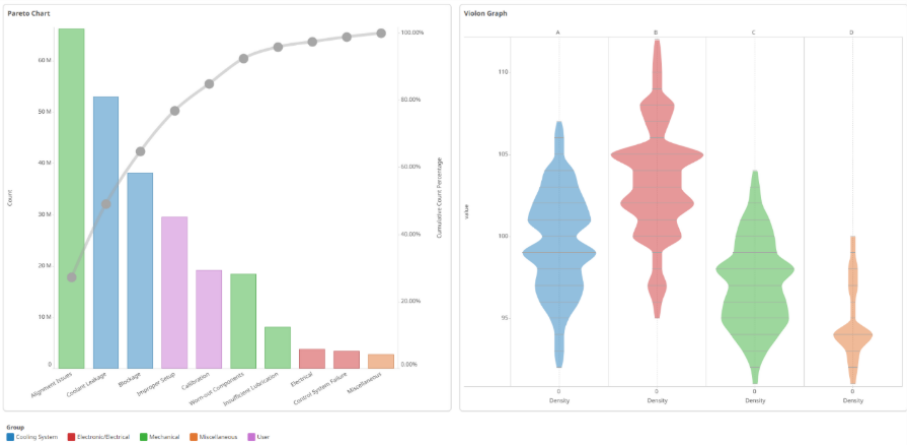

Figure 2 – Examples of the Pareto chart and Violon plot as important visualizations for RCA using Altair RapidMiner.

Altair RapidMiner offers comprehensive analysis capabilities that meet your needs and much more. You can access summary statistics and statistical process control, access powerful visualizations, assess data quality, perform data correction, and augment everything with machine learning and AI. Rather than being code-first, you can leverage a code-optional interface and build workflows with a visual interface; in other words, you only need to code if you want. And collaboration is built in, with users working seamlessly together in a shared, unified, real-time environment.

Step 3: AI

There are several key limitations associated with traditional RCA tools, often linked to the assumptions of many statistical methods. These methods incorrectly assume that the observations are independent (with no correlation between variables); that variables follow specific distributions; and that relationships are consistent through time. In practice, these assumptions often don't hold true, which can distort analysis.

This is where machine learning and AI can provide more complete, more impactful results. For example, we might use an interactive decision tree (shown below) to understand the different segments within our dataset. Within seconds, a dataset can be smartly segmented using machine learning algorithms, and we can see groups within the data and the features these groups have in common.

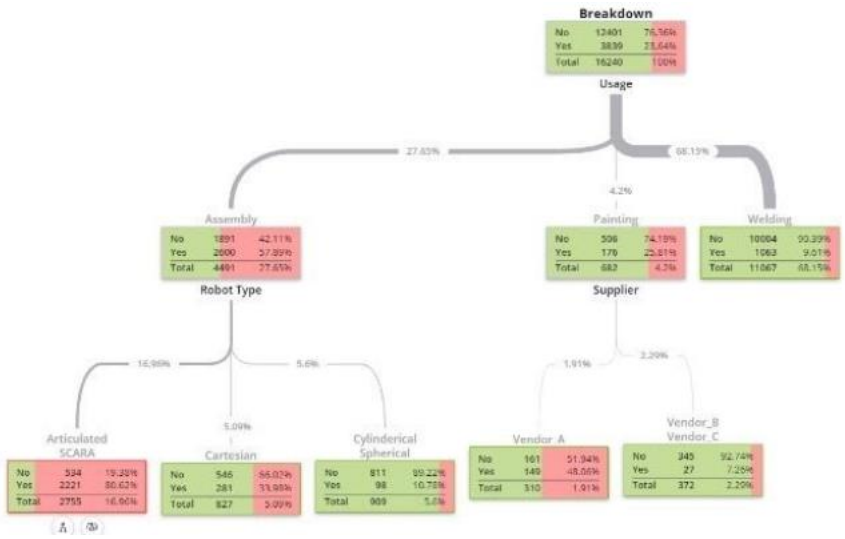

Figure 3 – Interactive decision tree for highlighting common features of failures across many variables.

Machine learning is also used in RCA to build predictive models that are trained to predict failures; using these models can detect patterns in the data as well. This allows organizations to shift from reactive analysis – powered by post-event data analysis and reports – to proactive decision-making. Using machine learning-powered RCA, teams can predict key metrics in real time, generate predictive alerts, and automatically recommend actions.

Altair RapidMiner empowers users to build these predictive models – but as an enterprise platform, its capabilities go much further. For example, once trained, users can deploy and manage predictive models with machine learning operations (MLOps), connect to operational technology (OT) and information technology (IT) with Internet of Things (IoT), and integrate both data and predictions into dashboards for immediate insight.

When considering generative AI applications, Altair RapidMiner can also introduce AI agents to support RCA. You can create autonomous AI-powered systems to prepare and analyze data independently, automatically find patterns and correlations within data, and have the agent take or direct corrective action without any additional prompting.

Conclusion

RCA is pivotal for organizations seeking to improve efficiency, enhance product quality, avoid production delays, and minimize risks. By leveraging advanced analytics, AI, and automation, businesses can transform RCA from a reactive problem-solving exercise into a proactive strategy for continuous improvement. Altair RapidMiner provides the tools needed to streamline data access, analysis, and AI-driven insights, enabling organizations to uncover hidden patterns and make data-driven decisions. As industries continue to evolve, embracing a more structured, more intelligent approach to RCA will be essential for staying competitive.

To learn more about Altair RapidMiner's RCA solutions, visit altair.com/manufacturing-analytics.