comet

SEE THE SIGNALS:

Using AI to spot leading indicators of environmental incidents

Turning complex data into actionable insights for environmental safety

A white paper by: Leith,D., Nicholas,P., Gowdar,P.N. and Ande,K.

Overview

Organisations in high risk industries, such as water supply and wastewater management, gather large volumes of structured and unstructured data and information about incidents concerning their assets. This ranges from customer reports of issues, through internal observations and near miss reports, to full incident reports. Together this represents a considerable resource to assist in decision making and understanding past, present and future risk.

Such organisations struggle to draw these sources together and extract meaningful intelligence. This can be due to lack of structure, or a common data structure; or systems unable to communicate together; or a lack of tools to analyse the data to extract actionable intelligence; or poor quality data and information.

Artificial intelligence (AI) tools, in particular natural language processing (NLP) and machine learning (ML), present an opportunity to make sense of this information allowing hypotheses to be tested, tracking trends, and pre-empting future incidents. The ultimate goal to achieve is therefore enhance decision making to better target investment and interventions.

This paper outlines development using the COMET® AI module to ingest large data sources to learn and train models which turn data into actionable intelligence. This approach is illustrated with a case study in the wastewater industry.

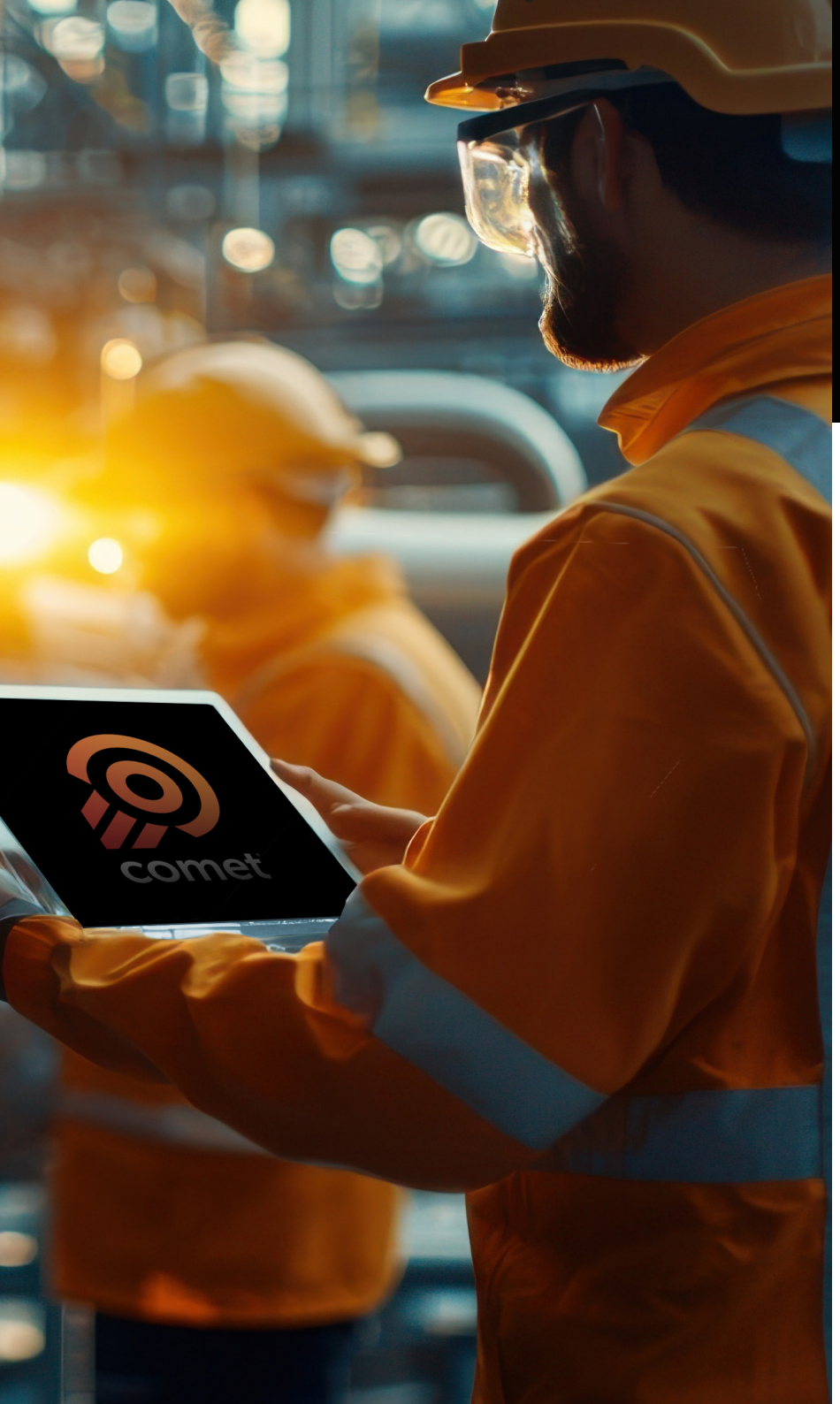


Table of CONTENTS

Introduction

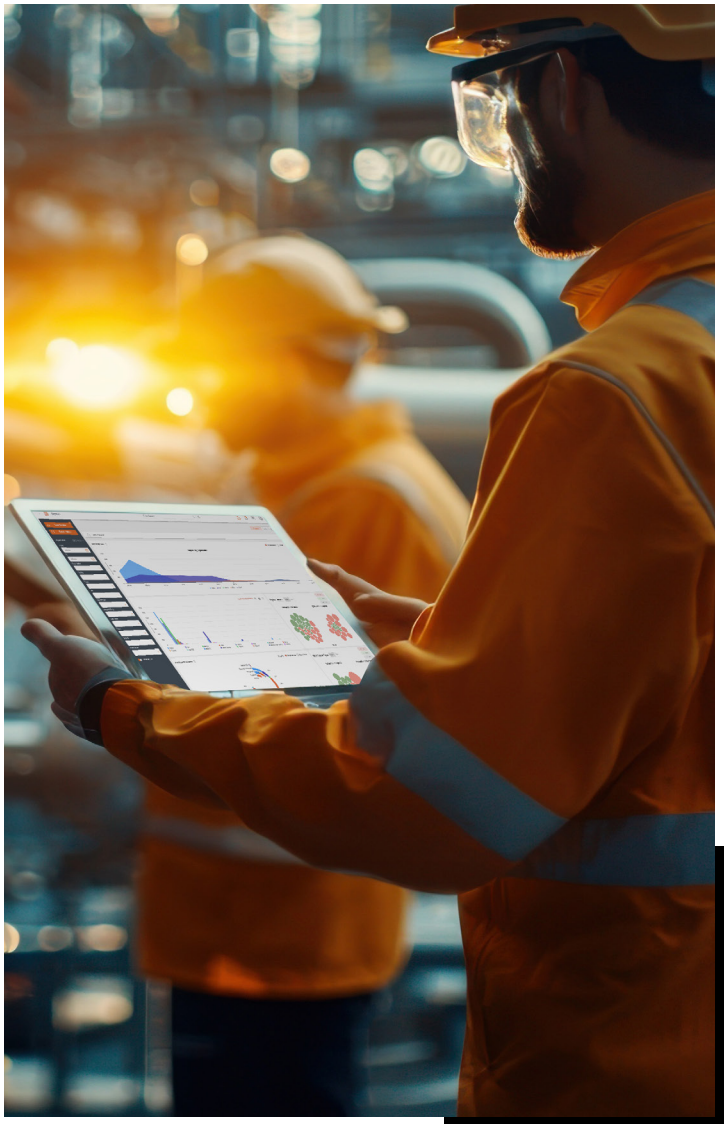
AI tools, such as NLP and ML, have the potential to be utilised to create advanced data analytics to drive down environmental incidents and improve asset performance through better informed investment decisions and the development of predictive analytics.

Data sets can be managed, ingested, trained and turned into actionable intelligence allowing better planning and prioritisation of capital investment.

Such tools present an opportunity to present an accurate and informative initial insight of systematic issues, hotspots and trends, derived from the free text held in incident reports, near misses and observations, therefore allowing preventive actions to be identified, applied and tracked.

When analysing an environmental incident, near miss or observation, some insight simply can't be determined using the human eye alone. AI has the ability to analyse tens of thousands of records to recognise industry specific terminology and extract unseen conclusions from the patterns that emerge in natural language. Findings can then be displayed, suggesting areas for improvement, and enabling professionals to be more proactive than reactive.

The authors have developed Natural Language Processing (NLP) capability within the COMET risk management software suite to perform analysis on the free text description of incidents or events and determines insights such as key recurring topics or phrases, commonly occurring hazards, severity, and typical root cause categorisation. It highlights not just a mythical primary cause, but multiple causes spanning the entire organisational spectrum.



These can be presented and correlated, displaying each category in order, and revealing trends. When combining these insights with some standard classification of the events, such as contractor/project, or location, we can start to build some powerful views of the data and address underlying issues, and using the date of event look at seasonal trends.

The models developed have the capability to read and classify each report, investigation or close call according to 4 criteria:

- Topics – which activities, equipment or environmental incidents are described in the reports.
- Hazard – which hazard, condition or contributory factor has occurred.
- Root Cause – which COMET root cause category can be correlated with the reports and filters.
- Severity – understanding Severity aids Prioritisation, highlights bad actors and provides alerts to trends.
- Ultimately this would lead to better decision making on where to target interventions and investment.

The challenge of decision-making

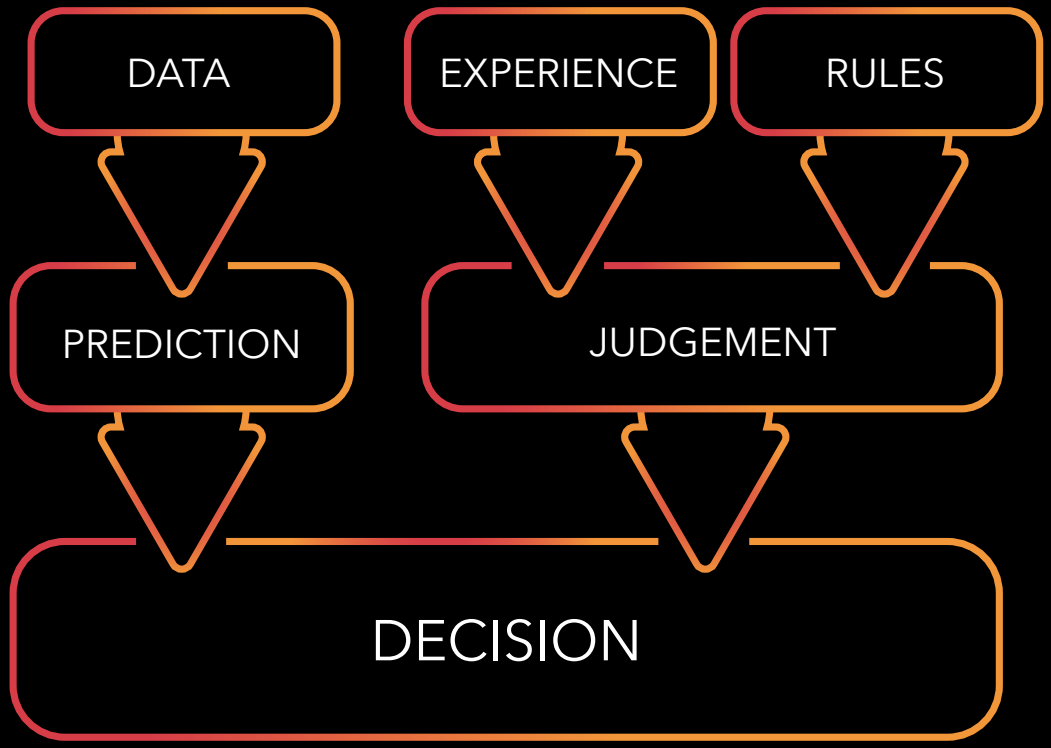
People make decisions using a combination of prediction and judgement. Judgement is based upon experience and any accepted rules.

Prediction is based upon data.

Some people avoid making decisions, or they make default decisions, because the cost or effort of getting data of enough quality to make a reliable prediction is too high. The result is that often organisations keep repeating the same ineffective actions, or direct investment in the wrong direction.

Therefore the role of data becomes a significant factor in whether or not risk is reduced. There is a strong driver to both utilise the data which does exist in order to extract risk intelligence, and there is an equally strong driver to improve the quality of the data which is being gathered and so maximise its value.

Figure 1 – Decision making components



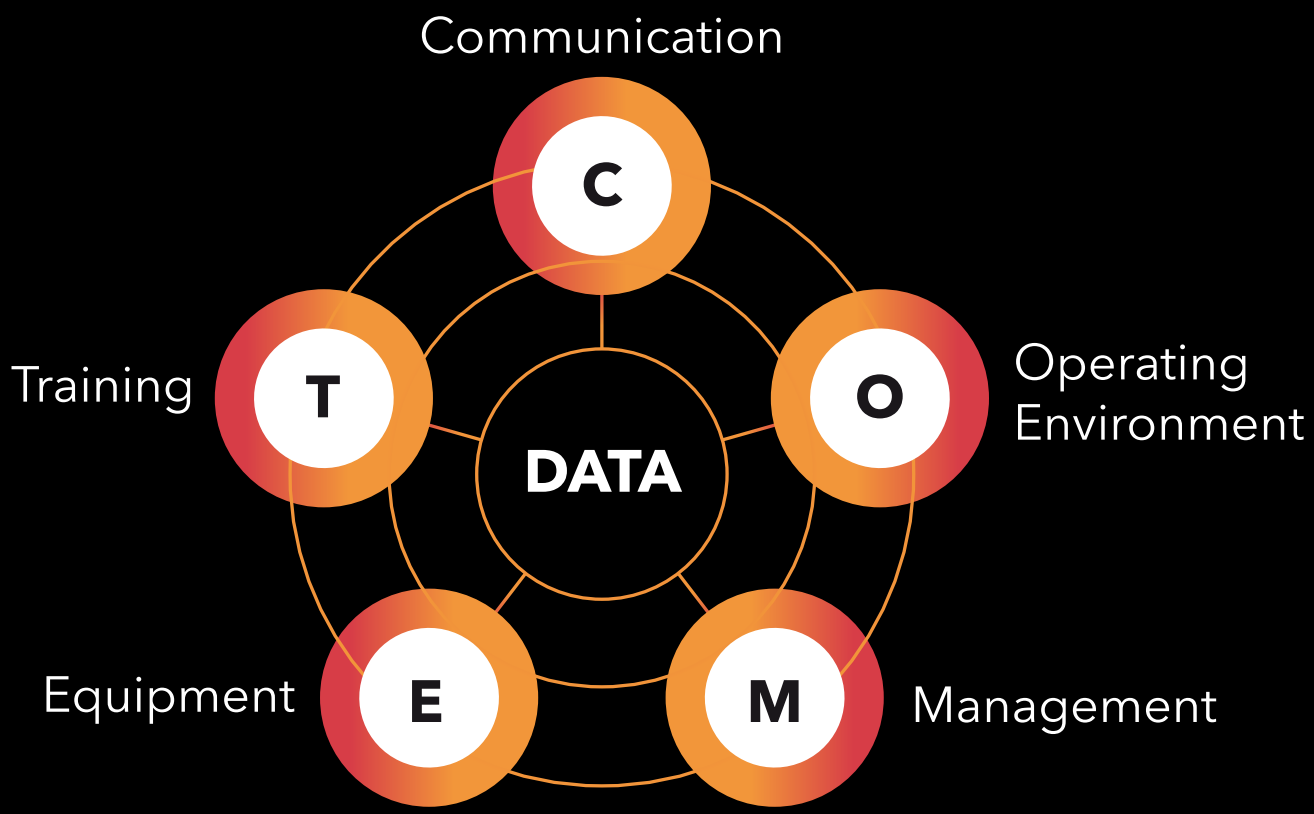
Connecting Data for Clear Insights

To address the issue of effective utilisation of unstructured data, the COMET AI tool was developed to 'listen at scale' - it deploys Machine Learning into large unstructured HSEQ datasets, enabling the identification of unseen insights in the data. It identifies topics within this data but importantly brings a structure to this previously unstructured data - it maps the ML derived inisghts back in to COMET's uniquely coded root cause analysis (RCA) taxonomy.

Intelligently mapping all diverse sources of data into this taxonomy allows connections to be made between unstructured and structured data which would have previously been impossible.

This approach was applied to a particular case study in the wastewater industry. The wastewater management organisation held diverse data sources on environmental spill incidents. This included not just obvious sources such as reports of spill incidents, but also extensive Customer Relationship Management (CRM) system records where customers have reported issues – these CRM records provide an extensive log of a description of the reported spill, the dispatching of the crew, the crew reports, the asset records, etc. No use was made of the data in the system beyond the unstructured log.

Figure 2 – COMET Root Cause Analysis taxonomy

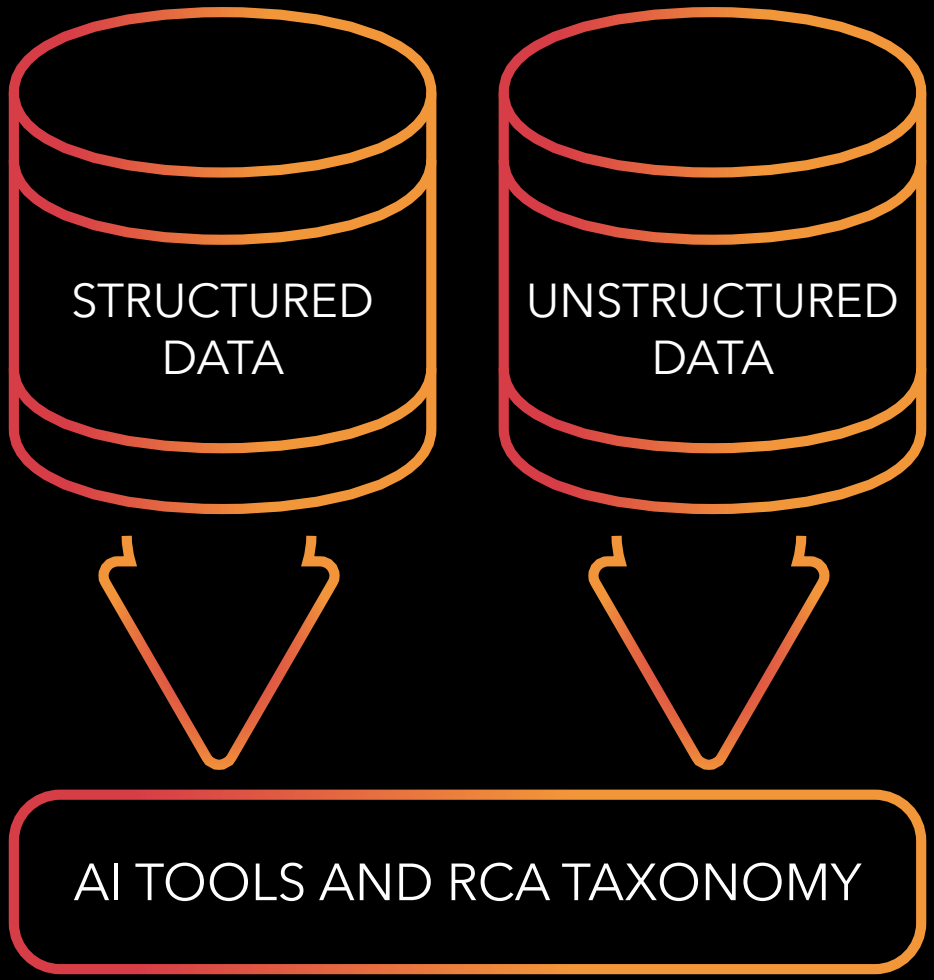


Data quality and making sense of the data

In applying this approach to a number of organisations it is clear that the quality of existing data on incidents is highly variable. Some organisations, or teams within organisations, produce detailed reports and keep extensive records. In our wastewater spill case study the CRM data had the potential to be useful but was disjointed and unstructured and no simple narrative was derived from each report logged. Other data sources were, however, more structured and of a more easily manipulated quality, for example classification of the type and severity of the spill incident, and whether it was a single or repeat incident.

Using the AI model and the RCA taxonomy provided the glue to bring these varied data types together. For example, the CRM logs which typically comprised of 500 words of unstructured information associated with a single flooding incident, had the context extracted using data science tools, including OpenAI, into a concise summarised overview of what happened. We can then start classification on the combined data sets, by using a data science technique called topic modelling, create 'topics'. Topics are multi word themes that emerge from the text in the data sources, due to their frequency and context. For example they may identify a topic as "flooding syphon cellar" indicating in this case that there is a strong theme to the occurrence of flooding issues involving syphon in cellar locations. These topics are the first level of abstraction to signal some risk intelligence which is contained in the data. The topics are pointers to explore further.

Figure 3 – Combining structured and un-structured data sets



Visualising intelligence and testing hypotheses

With diverse data sets processed by the AI tools it then becomes possible to visualise the aggregated data and the intelligence extracted into a single dashboard. This allows an organisation to have a holistic view of data it previously was unable to see.

Such visualisation includes simple views of the number of data points relating to particular topics. For example finding all reports of incidents, observations or activity associated with a blockage or flooding involving a syphon.

The visualisations from the model also go further and show the results of learning the type of root causes as deduced from the data. For example, Figure 5 shows that the most common root cause categories for the wastewater flood incidents relates to Operating Environment causes and, less but also significantly, Management causes.

Figure 4 – Visualisation and drill-down into specific themes in diverse data sets

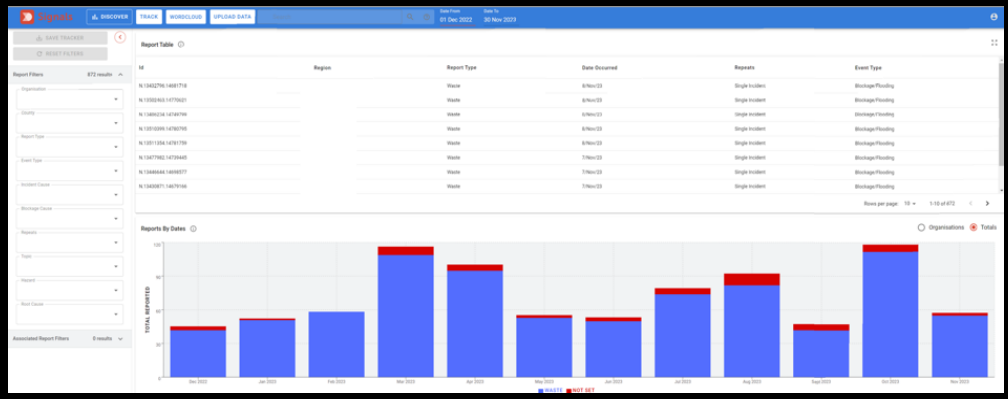
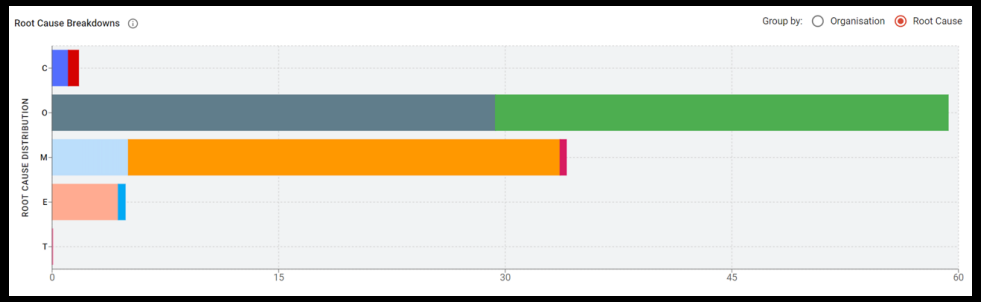


Figure 5 – Learning of root causes of an incident type



These types of incidents are significant as they challenge perhaps assumed thinking that the problem is with the equipment, when in actual fact the underlying problem is with the situation in which the equipment is operated and the management decisions around the incidents.

These root causes can be explored further in the tool to drill into specific incidents and understand what was going on – what was effective and what was ineffective. It also starts to highlight systemic issues if the same root causes are behind the same incident types.

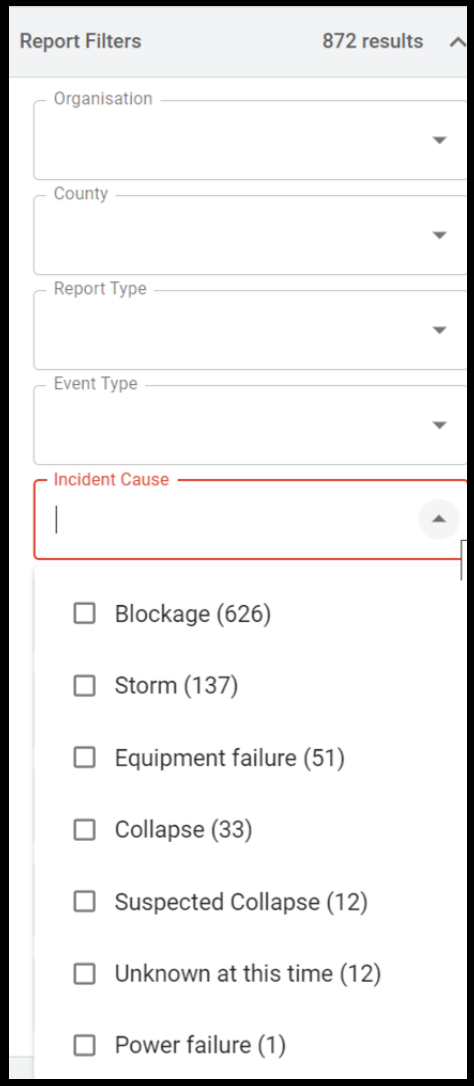


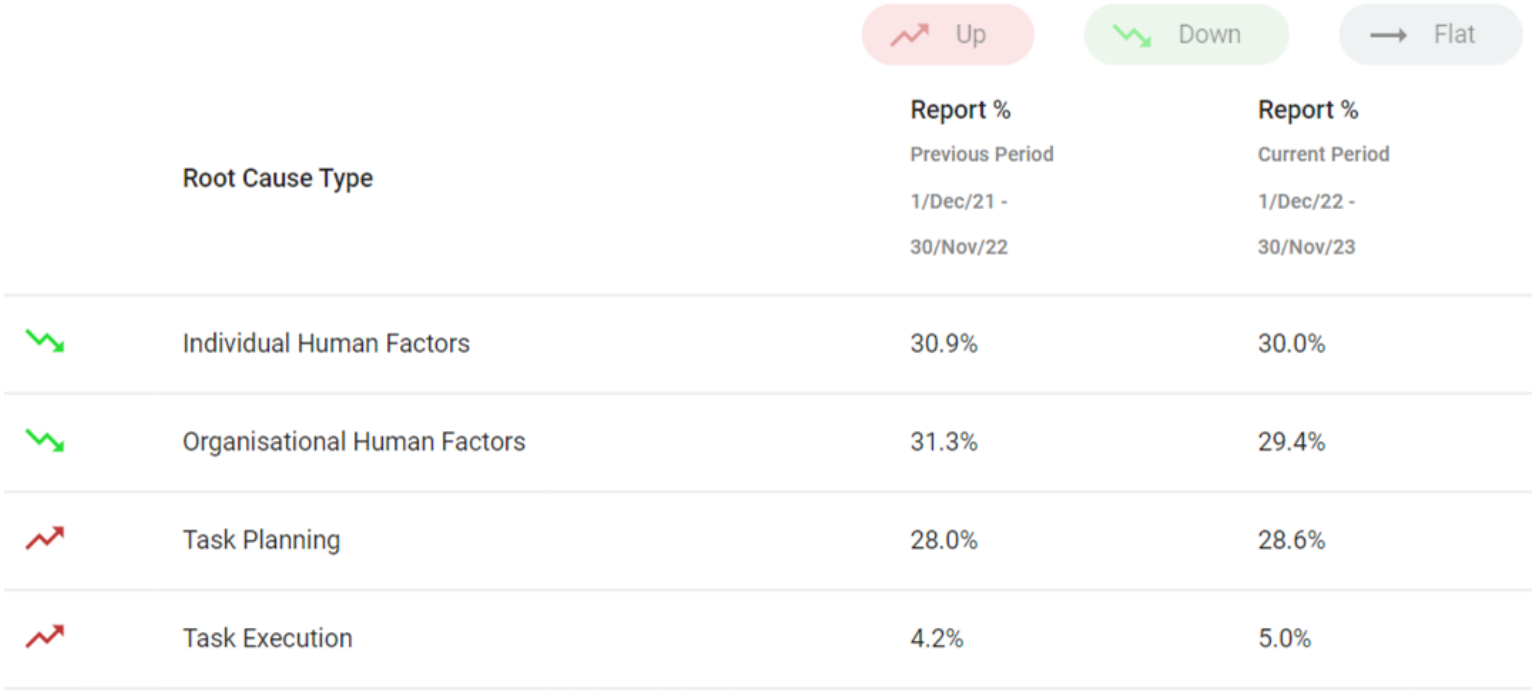
Figure 6 – Learned data structure within filters

Equally, this tool can be used to test hypotheses – if an organisation suspects there may be certain correlations or themes occurring then it can select from filters which have been generated by the machine learning model.

Tracking changes and trends for reducing risk

The AI model outputs can be interrogated across time periods to explore trends in topics or root causes. For example Figure 7 shows that root causes associated with Organisational Human Factors have demonstrated an improvement in comparable years.

These items can also be set with alert thresholds so that as the model processes further new data and the trends change then an organisation can be alerted to potential future risk.



	Root Cause Type	Report % Previous Period 1/Dec/21 - 30/Nov/22	Report % Current Period 1/Dec/22 - 30/Nov/23
↘	Individual Human Factors	30.9%	30.0%
↘	Organisational Human Factors	31.3%	29.4%
↗	Task Planning	28.0%	28.6%
↗	Task Execution	4.2%	5.0%

Figure 7 – Selecting elements for trends and alerts

Making better decisions with AI

With the output from this tool an organisation is able to better make decisions on a number of fronts:

- **Decide on most appropriate interventions, for example training for maintenance crews; improvements to operating procedures; etc.**
- **Decide on best targeted investment spend, for example spending capex on improvements to cellar layouts thus preventing future risk rather than on dealing with the aftermath of floods.**
- **Decide on optimised resource allocation, for example deploying people on tasks which make a difference to reducing future risk.**
- **Decide on pre-emptive actions, for example tackling latent factors which are increasing over time and likely to become future incident root causes.**
- **Decide on improvements to data gathering, for example structuring the way in which in incident data is recorded in a common taxonomy.**

Conclusion

The work described in this paper has demonstrated that it is possible for the wastewater industry to harness AI technology to analyse tens of thousands of unstructured and structured records relating to environmental incidents to extract unseen conclusions from the patterns that emerge in natural language. Findings can then be displayed, suggesting areas for improvement, and enabling professionals to be more proactive than reactive.

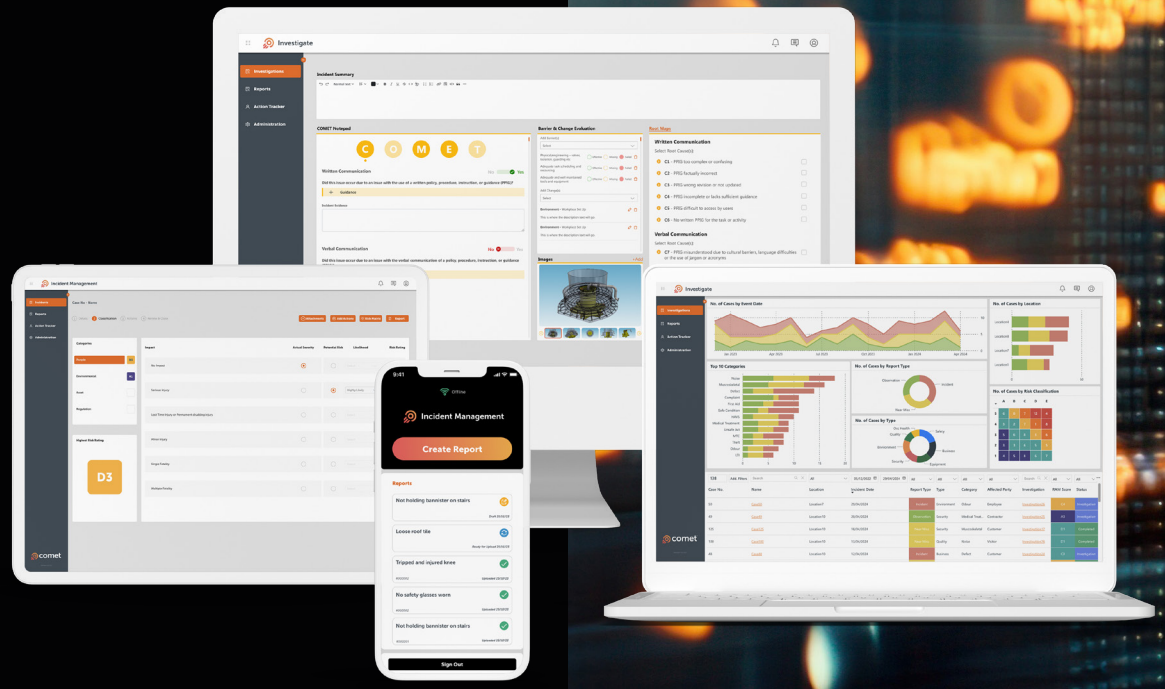
While it has been demonstrated that such a tool can work with unstructured data, organisations still need to drive to improve the quality of the data which they gather to maximise the benefits which can be derived from using AI, and so improve decision making on interventions and investment.

Taking action with COMET

Harness AI to proactively manage risks and prevent incidents

In high-risk industries, managing vast amounts of Health, Safety, and Environment (HSE) data can be a significant challenge.

Traditional methods often fall short in extracting meaningful insights from diverse and unstructured data sources. COMET Signals addresses this by using Artificial Intelligence (AI) and Machine Learning (ML) to transform complex data into actionable intelligence.



Key Features of COMET Signals:

IN AI-DRIVEN ANALYSIS:

Leverages AI and ML to process extensive HSE datasets, uncovering hidden insights and root causes of incidents.

REAL-TIME ALERTS:

Provides immediate, essential information through clear visualisations, enabling timely interventions to prevent failures.

SEAMLESS INTEGRATION:

Fully compatible with the COMET suite, including optional modules for audits, investigations,and supply chain management.

By adopting COMET Signals, organisations can move from reactive to proactive risk management, improving safety performance and operational efficiency. AI-driven insights enable informed decision-making, targeted interventions, and continuous safety improvement.

Discover how COMET Signals can transform your approach to risk intelligence

"GET IN TOUCH TODAY"

+44 0 (1224) 937479

info@cometanalysis.com

www.cometanalysis.com