Using AI and ML to Analyze Incident Reports

Dr. Fereshteh Sattari, Dr. Renato Macciotta, Dr. Lianne Lefsrud, & Daniel Kurian

Fall, 2021

Research Problems

- While companies keep incident data in 1000s of reports, rarely do they analyze these to learn and prevent future incidents.
- Further, related datasets (maintenance data, performance data, employee survey data) are rarely integrated to understand these as leading indicators.

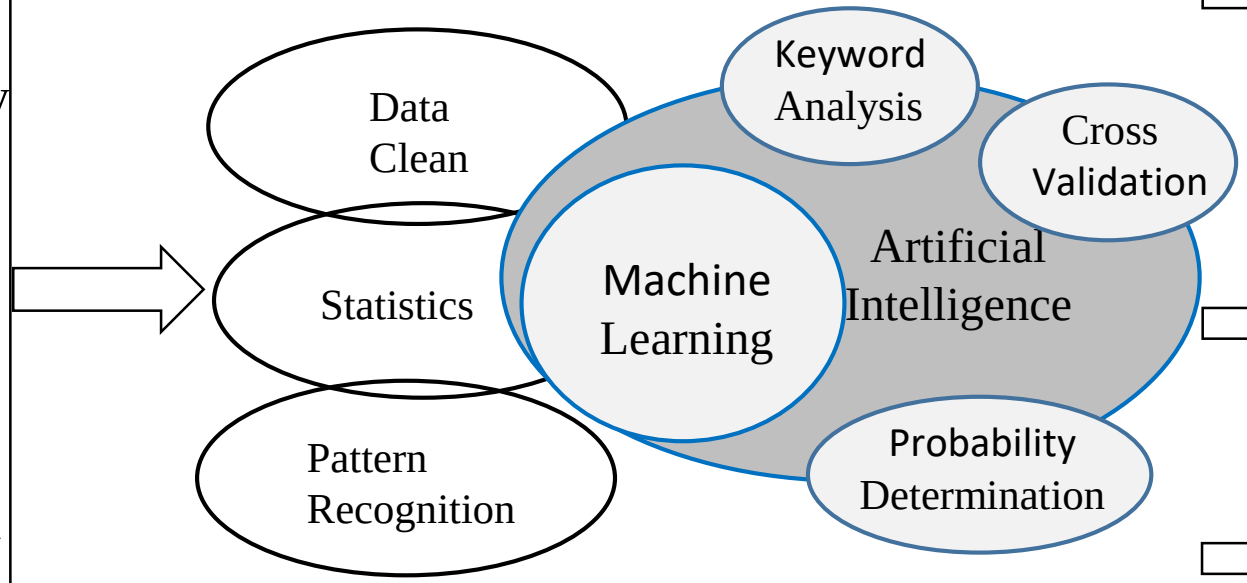
Objectives

- Use Machine Learning (ML) and Keyword analysis (KA) to improve consistency and accuracy when evaluating risk, and deliver practical and tailored outputs to prevent and mitigate risk.
- Identify which Asset Integrity Management (AIM) elements most influence the occurrence of an incident/accident rate, by using AI/ML (Bayesian Network Analysis, Cross-correlations, etc.).
- Integrate a quantitative data set with complementary datasets (interview, survey data) and provide further explanatory analysis.
- Provide practical recommendations to reduce/eliminate these latent causes.

Research Methodology

Data Required

- Incident report
- Perception survey
- Standard operating procedures
- Performance data



Results

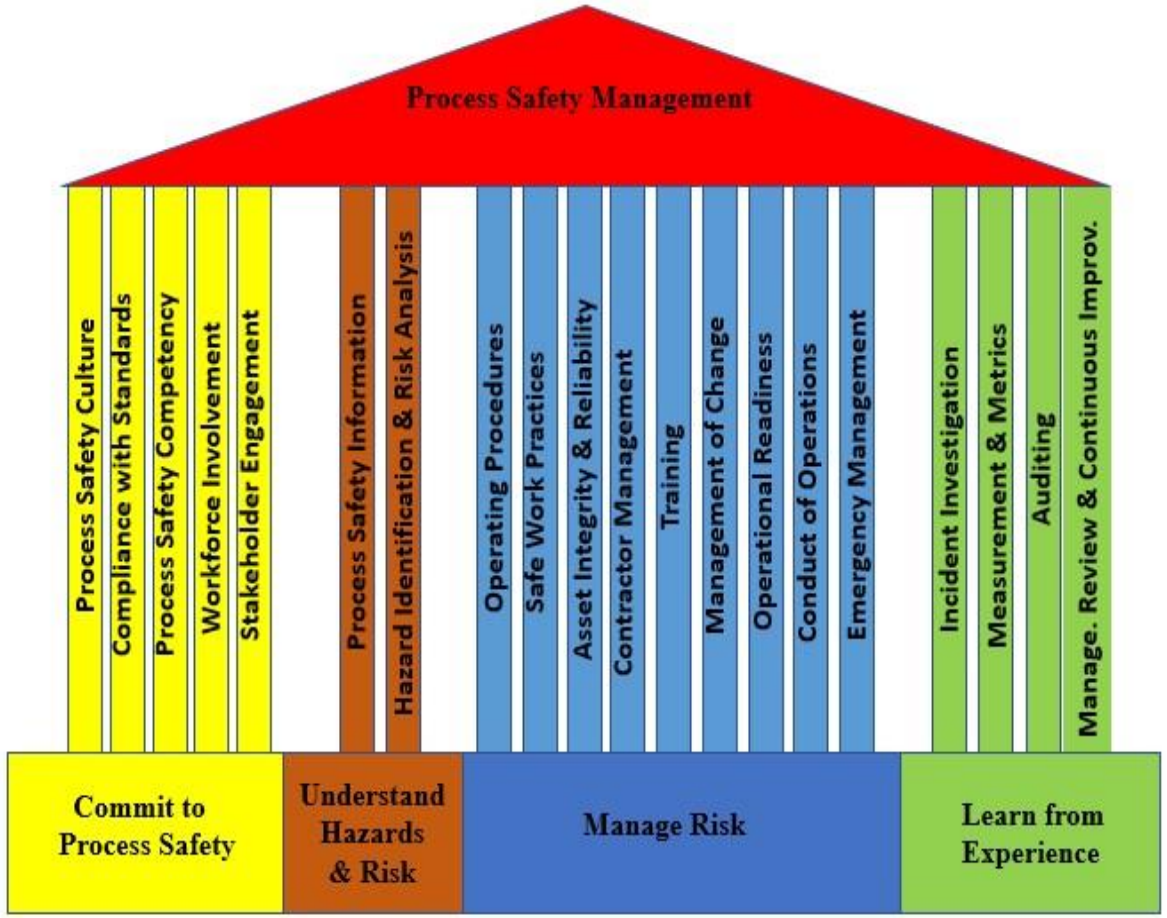
- Automatically "clean" databases to categorize incident types.
- Consistent method for reporting incidents and evaluating risks by assessing frequency and describing actual and potential consequences.
- Analyze the relative influence of and relationship between various leading indicators to discover patterns, identify best leading indicators, and predict incidents.
- Produce more accurate results by employing a fusion of both qualitative and quantitate data sources.

Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Categorize Incidents into PSM Elements by the Application of ML & KA



PSM Element	Keyword
Compliance with Standards	compliance, comply, regulatory
Process Safety Information	psi
Hazard Identification & Risk Analysis	hazard, risk
Operating Procedures	procedure
Safe Work Practices	safety, safe work
Asset Integrity & Reliability	equipment, asset
Contractor Management	contractor, third party, 3rd party
Training	training, inexperience, lack of experience
Management Review & Continuous Improvement	management

Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Bayesian Network Analysis (BNA)

BNA: this technique provides a qualitative description of relationships among variables of interest in a large data set and provides an optimal solution if the same result obtained from two or more methods. Bay's rule:

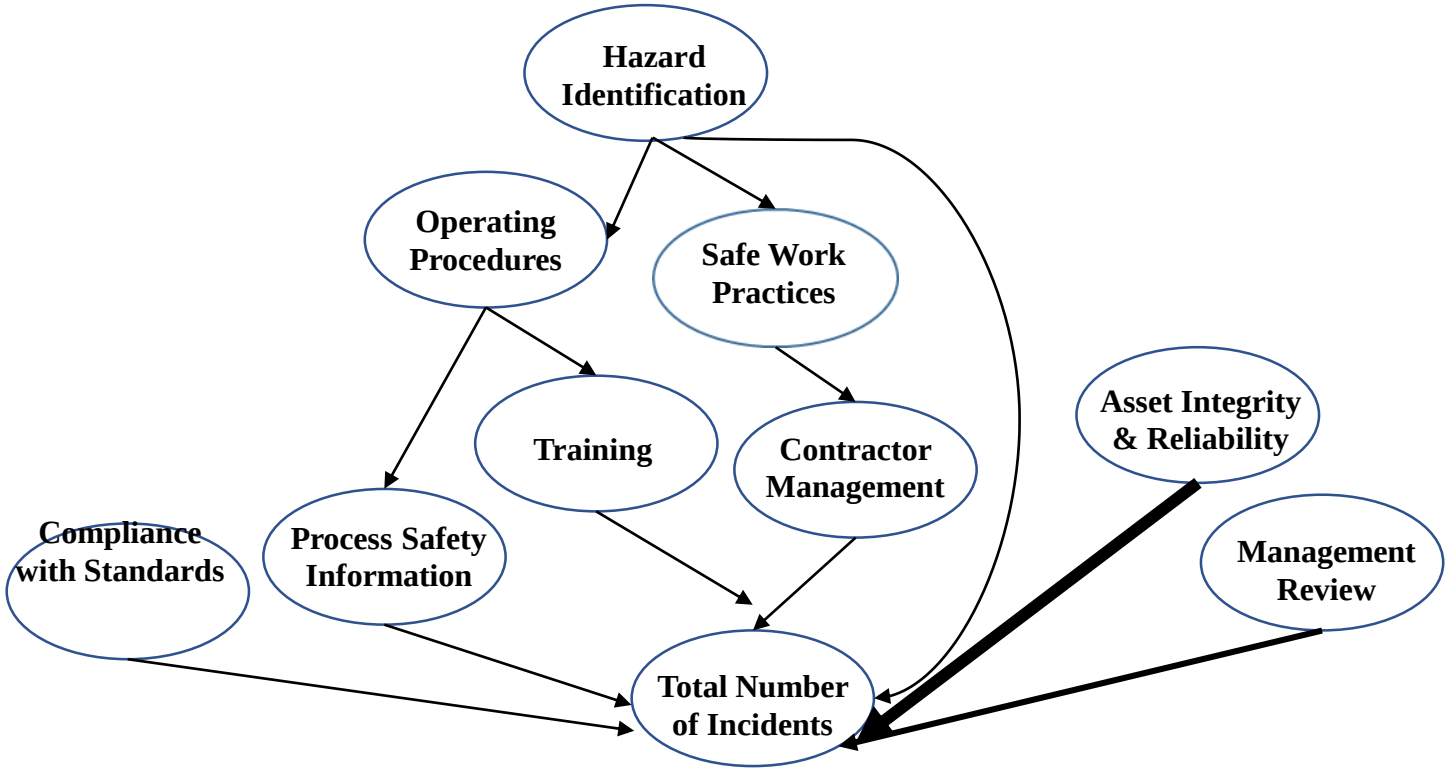
$$\mathbf{P(G|D) = \frac{P(D|G)\,P(G)}{P(D)}}$$

P(G|D) is the posterior distribution of the parameter; P(D|G) is the likelihood function; ; P(G) is the prior distribution of the parameter; and P(D) is the marginal distribution of D or constant

1. **Hill-climbing (HC):** it starts with an initial guess of a solution, then iteratively makes local changes to find the best possible solution.

2. **Tabu:** it uses a memory list to guide the search, starting with a feasible initial solution and picking the next best option that can increase the score function.

Develop the Map between PSM Elements and the Total Number of Incidents



From Group	To Group	Arc Strength
Asset Integrity & Reliability	Total Number of Incidents	-315.06
Management Review & Continuous Improvement	Total Number of Incidents	-102.35
Hazard Identification	Total Number of Incidents)	-78.86
Contractor Management	Total Number of Incidents	-52.13
Operating Procedures	Total Number of Incidents	-50.86
Safe Work Practices	Total Number of Incidents	-30.81
Training	Total Number of Incidents	-9.96
Process Safety Information	Total Number of Incidents	-4.73
Compliance with Standards	Total Number of Incidents	-3.49
Hazard Identification	Operating Procedures	-7.84
Hazard Identification	Safe Work Practices	-4.63
Operating Procedures	Training	-1.67
Safe Work Practices	Contractor Management	-1.07

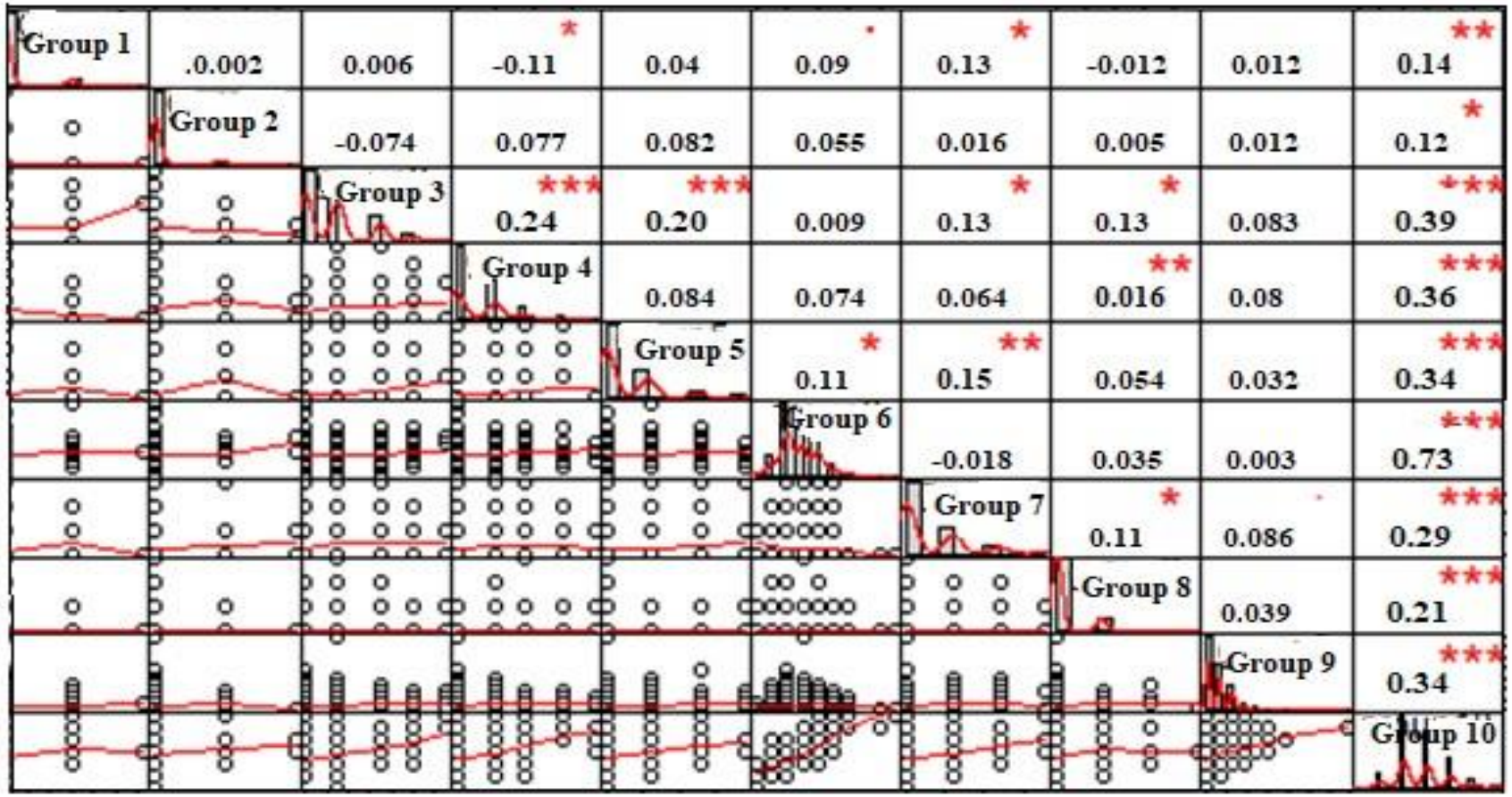
Note: Total number of incidents has maximum dependency with *Asset Integrity & Reliability*; by addressing this, we can reduce the total number of incidents by half.

Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Result Validation by Linear Cross Correlation Method

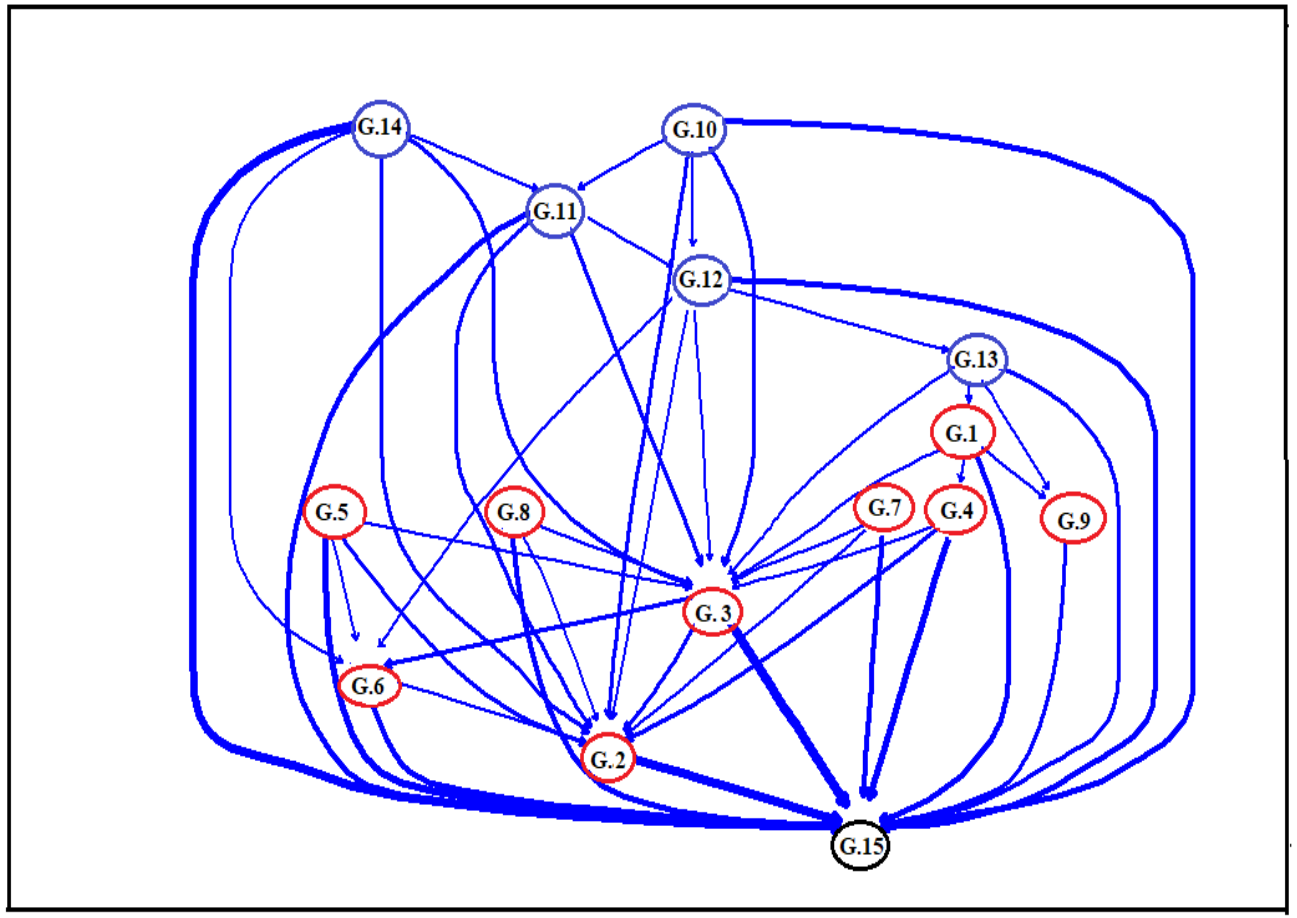


Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Map between Asset Integrity Management Elements and of Incidents



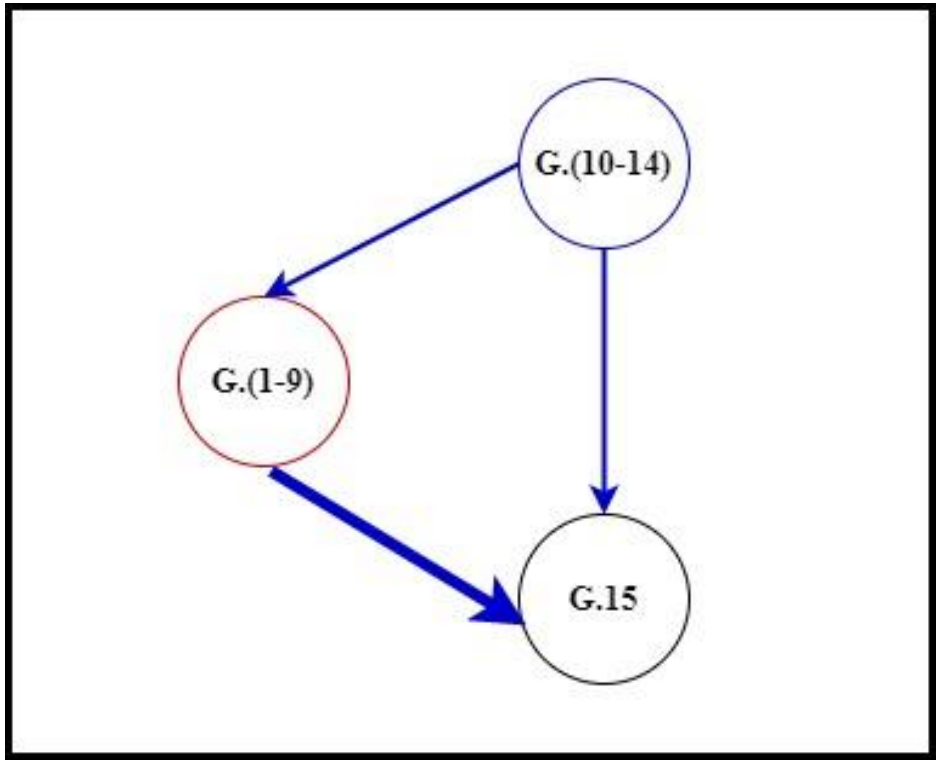
Asset Integrity Management (AIM)			
Asset and Procedure (Element # & name)		Human Factors (Element # & name)	
G.1	Design	**G.10**	Management understanding and support
G.2	Construction	**G.11**	Communications across the lifecycle stages
G.3	Testing	**G.12**	Training- Sufficient understanding of the design and construction features
G.4	Operation	**G.13**	Contractors/ external specialist resources
G.5	Inspection/Audit	**G.14**	Human error
G.6	Maintenance/Obsolescence of the equipment		
G.7	Mechanical equipment		
G.8	Safety-critical protective systems/Control system		
G.9	Electrical/ Instrumentation		

Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Revealing the Importance of Human Factors on Incident Rates



From	To	Arc strength	Dependency with the incident rate
Asset	Total number of incidents	-564.39	65.32%
Human factor	Total number of incidents	-203.76	23.58%
Human factor	Asset	-95.9	11.10%

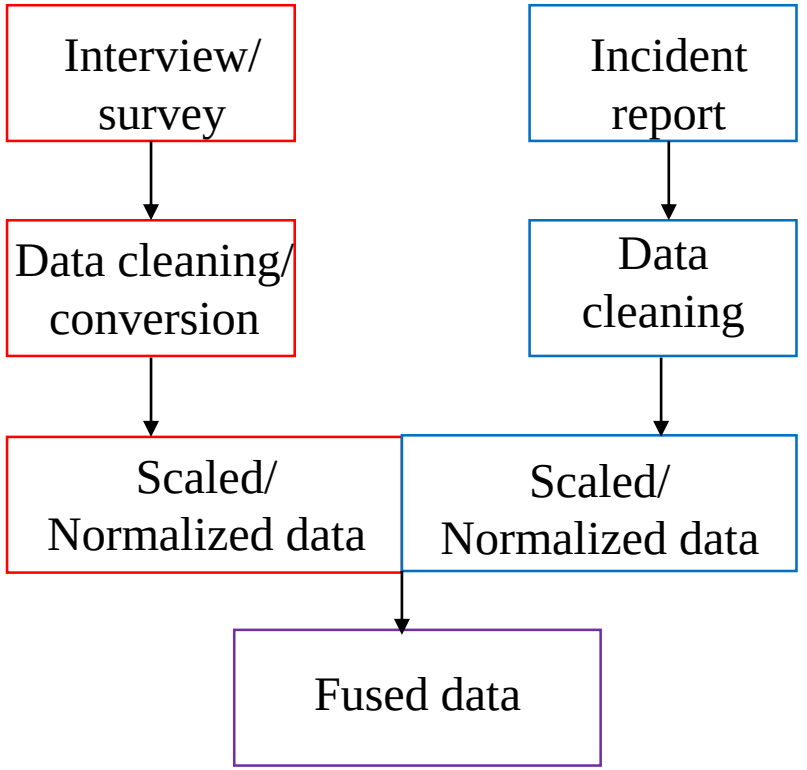
Note: Aside from direct asset or human factors causing an incident, it is also possible for human factors to cause asset or procedure failures.

Data Fusion

Data fusion methods have a long history of development starting with applications in robotics. Current application areas include mine detection, maintenance engineering, and weather prediction.

Proposed Methodology:

- Collect information/data from multiple sources of data.
- Employ a combination technique for fusion and concatenation of the data vectors from different sources into a single data vector for each data point.



Hazards31

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE

Conclusions

By analyzing incident reports we were able to:

- Implement ML and KA in order to categorize incidents, analyze risks, provide prevention and mitigation strategies, and identify the leading indicators.
- Apply BN to determine the importance of each contributing factor in accident rate, prioritize them, and map the way they are linked.
- Apply DF in order to combine one qualitative approach (conducting interviews) with the completely quantitative approach (AI) to improve the accuracy of the result.
- Recommend some actionable strategies to improve the safety practices.

Thank You for Listening!

Dr. Fereshteh Sattari, Sattari@ualberta.ca

Hazards31

IChemE Safety Centre

IChemE ADVANCING CHEMICAL ENGINEERING WORLDWIDE