CHEMICAL ENGINEERING TRANSACTIONS
VOL. 105, 2023

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: David Bogle, Flavio Manenti, Piero Salatino

ISBN 979-12-81206-04-5; **ISSN** 2283-9216

DOI: 10.3303/CET23105084

Unveiling the Achille's heel: Detecting Organizational Weaknesses in the Energetic Transition Challenge

Tomaso Vairo[a*], Patrizia Agnello [b], Fabio Currò [a], Bruno Fabiano [a]

[a] DICCA - Civil, Chemical and Environmental Engineering Dept., Genoa University, via Opera Pia 15, 16145, Genoa, Italy
[b] Department of Technological Innovations and Safety of Plants, Products and Anthropic Settlements, INAIL (Italian National Institute for Insurance Against Accidents at Work), Rome, Italy.
tomaso.vairo@edu.unige.it

To be able to thrive in the grand challenges of the current historical moment, which includes important driving phenomena such as climate change, digitalization and energy transition, the organizations need a comprehensive understanding of the organizational and technical aspects that may pose opportunities and risks. This paper presents a novel approach to identify weak organizational and technical factors within the context of the energetic transition challenge. To accomplish this, a Machine Learning system is proposed, that integrates, as input features, escalation and mitigation factors related to the risks that may arise in relation to the energetic transition. The target variable is an indicator concerning the possible increase in the probability of accidents and near misses, which is selected as an effective detector of potential weaknesses in the system. The primary objective is to uncover organizational aspects that influence the mitigation, or enhancement of technological risks during the energetic transition. By analysing the interplay between organizational and technical factors and their role on preventive and mitigating barriers, this paper aims at identifying critical areas that require Attention and improvement to ensure a smooth and successful energetic transition process. A reference case-study is presented to demonstrate the actual capability of the presented framework. The findings of this study have practical implications in the definition of organizational priorities in managing the energetic transition; the identified weaknesses can serve as a basis for targeted interventions and strategic decision-making, allowing for more effective risk management and improved outcomes during the energy transition.

1. Introduction

The global energy landscape is undergoing a significant transformation as countries strive to transition towards more sustainable and low-carbon sources of energy (Schischke et al., 2023). This energetic transition, driven by the need to mitigate climate change and reduce greenhouse gas emissions, presents new opportunities and challenges for various industries (Pasman et al., 2023). In this regard, new approaches both conceptual and combined data-experience driven, as well as innovative technical/engineering solutions are needed to achieve UN Sustainable Development Goals 2 (zero hunger), 3 (health and wellbeing), 6 (clean water and sanitation), 7 (affordable and clean energy), 9 (industry, innovation and infrastructure) and 12 (responsible production and consumption). While the transition to renewable energy sources brings numerous benefits, such as reduced environmental impact and improved energy efficiency, it also introduces new and emerging risks that need to be effectively managed (Chen et al., 2023). A peculiar aspect is related to challenges are still to be addressed in order to make further progress in safety and sustainability of plants and processes moving towards Safety 4.0, identifying the benefits of digitalization technology to the management of major hazards and ensuring that the risks of the technology are understood and correctly managed. Novel and improved methods for early warning, prediction of actual risk and safety warranty, both from a design point of view and in the dynamic situation of operation are currently investigated and teste at different scales (Pasman and Fabiano, 2021). An integrated approach implementing process and risk analysis aprioristically at the conceptualization stage of a process and to compare quantitatively emerging technologies in terms of intrinsic safety based on the Fire and Explosion Index (F&EI) allows the dynamic quantitative comparison of different design solutions (Bassani et al., 2023). In this study, the focus is the understanding the factors influencing safety, in terms of accidents and near

Paper Received: 15 February 2023; Revised: 3 June 2023; Accepted: 8 July 2023
Please cite this article as: Vairo T., Agnello P., Currò F., Fabiano B., 2023, Unveiling the Achilles' Heel: Detecting Organizational Weaknesses in the Energetic Transition Challenge, Chemical Engineering Transactions, 105, 499-504 DOI:10.3303/CET23105084

misses, within the context of the ongoing energetic transition. Based on accident and literature reviews and expert opinions, starting from the major contributing factors among leadership and safety culture, risk awareness, knowledge and competence, communication, and information and decision-making processes a checklist approach based on the Relative Efficiency Indicator (REI) allows evaluating the level of commitment of the top leaders in process safety management (Markowski et al., 2021). As industries shift towards new energy technologies and substances, new operational risks, technological complexities and workforce dynamics arise (Vairo et al., 2023). New technologies and processes entail new process and personal hazards, and that much effort is going into renewal towards energy transition, but safety analyses still scarce (Pasman et al., 2023). It is therefore crucial to identify and comprehend these factors to ensure the safety, protect the environment, and maintain the reliability and resilience of energy systems. In this study, a comprehensive approach to analyze and assess the emerging risks related to energetic transition is employed. The developed analysis is based on a combination of quantitative data and qualitative insights obtained through a questionnaire. The first set of features, representing various factors, is organized in a fishbone diagram. The fishbone diagram provides a systematic framework for categorizing and exploring the different aspects that contribute to the identified risks. Additionally, advanced machine learning techniques to model and predict the occurrence of accidents and near misses, which are selected as key indicators of emerging risks within the context of energetic transition. By leveraging a regression model, we aim to identify the influential factors and subfactors that contribute to these incidents, thereby providing valuable insights for risk assessment and mitigation strategies.
The primary objectives of this study are to:

- Analyze the factors and subfactors that contribute to emerging risks in the energetic transition.
- Develop a regression model to predict the occurrence of incidents and quasi-incidents.
- Identify the influential factors and subfactors that play a significant role in the occurrence of these incidents.
- Provide actionable insights for risk assessment and effective decision-making in the context of energetic transition.

By gaining a deeper understanding of the factors and subfactors influencing emerging risks, we can proactively address and manage these risks, ensuring a smoother and more sustainable transition towards cleaner and more efficient energy systems. To achieve this, a Random Forest Regressor (Breiman, 2001) is implemented, a powerful machine learning algorithm capable of capturing complex relationships and providing insights into feature importance. The features include information on hazard levels, contextual aspects, technical considerations, human and organizational factors, as well as anomalies and non-compliance issues specific to the organizational aspects of safety management. As recently commented (Pasman et al., 2023), human and organizational failures, failing safety management system, and poor safety culture play in this an important role of which the effects are not easily predictable. So, we should dedicate effort to consider safety aspects, means to control risks, introduce preventative and protective risk reducing measures. On these grounds, the focus is to uncover the key drivers of safety accidents in this evolving energy landscape, according to a two-step methodology focusing on:

- developing a predictive model using the Random Forest Regressor to forecast safety accidents and near misses based on the identified factors in the context of the energetic transition;
- providing a detailed analysis of feature importance, shedding light on the factors that have the most significant impact on safety outcomes in this changing energy sector.

By achieving these objectives, organizations could be able to proactively address emerging risks associated with the energetic transition and ensuring the safety of their operations.

2. Materials and Methods

2.1 Data analysis and preprocessing

The dataset consisted firstly of factors and subfactors, related to the conditions of the corporate organization towards the challenges posed by the energy transition, as assessed by different hierarchical levels within companies. Secondly, the dataset included the validated data retrieved from reports collected by the organizations, according to the safety management systems and covering the following items: occurrence of incidents, near misses, anomalies, and on-conformities assessment.
The factors were considered according to the following classification:
Accelerating factors

1. *Per*: This factor represents the overall hazard level and was further divided into subfactors:
    a. *Sost* (Substances characteristics),
    b. *Proc* (Process characteristics),
    c. *Imp_mat* (Plant and materials characteristics).

2. *Cont*: This factor represents the contextual aspects of the system and was further divided into subfactors:
    a. *Log* (Logistic),
    b. Interf (Interferences),
    c. *Soc_pol_fin* (Societal, Political and financial aspects).

Mitigating factors

3. *Tec*: This factor accounts for the technical aspects of the system and was further divided into subfactors:
    a. *Affid* (Reliability)
    b. *Innov* (Innovation).
4. *Um_org*: This factor represents the human and organizational aspects and was further divided into subfactors:
    a. *Con* (representative for the overall level of knowledge),
    b. *Cul* (Cultural aspects),
    c. *Comun* (Relational aspects).

Operative experience factors

5. *Inc*: This represents the occurrence of accidents.
6. *Q_inc*: This item represents the occurrence of near misses.
7. *Anom*: This item represents anomalies occurrences.
8. *NC_maj*: This factor represents major non-conformities.
9. *NC_min*: This factor represents minor non-conformities.

The factors under heading 1 and 2 are considered as accelerating factors for the energy transition risks, while the factors 3 and 4 are the mitigating factors. These data are all included in the model as categorical variables. Concerning the second set of factors (i.e., 5, 6, 7, 8, 9), it should be remarked that they are included in the model by deriving appropriate prior probability distributions from the collected data:

- *Inc*: Bounded-normal distribution with an expected value of 0.
- *Q_inc*: Bounded-normal distribution with an expected value of 1.
- *Anom*: Bounded-normal distribution with an expected value of 5.
- *NC_maj*: Uniform distribution between 0 and 100.
- NC_min: Uniform distribution between 0 and 100.

The whole data set was subsequently preprocessed to prepare it for model training. The categorical features were encoded using label encoding (Shah et al., 2022) to convert them into numerical values. This encoding step aims at ensuring that machine learning algorithms can properly interpret the data.

2.2 The regression model

The Random Forest Regressor (Breiman, 2001) is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree in the random forest is trained on a random subset of the data and features, which helps to reduce overfitting and improve generalization.
Random Forests construct many individual decision trees at training. Predictions from all trees are pooled to make the final prediction. The Random Forest Regressor was chosen as the machine learning model for this study due to its ability to handle both numerical and categorical features, as well as its capability to capture complex relationships within the data. The dataset was split into training and testing sets using an 80:20 ratio.
Let's denote the training dataset as *X_train*, which consists of n samples (X_train = {x1, x2, ..., xn}) and their corresponding target variables *Y_train* (Y_train = {y1, y2, ..., yn}). According to the Random Forest Regressor, a collection of decision trees {T1, T2, …, Tm} is built, where each decision tree Ti is constructed as follows:

1. Randomly select a bootstrap sample of the training data, denoted as Xi, by randomly sampling n examples with replacement from *X_train*.
2. Randomly select a subset of features, denoted as Fi, by randomly choosing a fixed number of features from the total feature set.
3. Train a decision tree Ti on the bootstrap sample Xi using the selected features, Fi. At each node of the decision tree, the best split is determined based on minimizing the mean squared error (MSE) criterion.

Steps 1-3 are repeated until a complete collection of m decision trees is developed.
To make a prediction for a new input sample x, the Random Forest Regressor aggregates the predictions of all decision trees in the forest. The final prediction is computed as the average of the predicted values from each tree. The Random Forest Regressor was trained on the training set using 1000 decision trees.
The trained regressor was then evaluated on the testing set to assess its performance in predicting the target variables (*Inc* and *Q_inc*). During the training process, the Random Forest Regressor was also utilized to assess the *feature importance* (Gerstorfer et al., 2023), which indicates the relative importance of each feature in

predicting the target variable. The feature importance is calculated based on the total reduction of the mean squared error (MSE) caused by each feature across all decision trees in the forest. To evaluate the accuracy performance of the Random Forest Regressor, the Root Mean Squared Error (RMSE) is calculated as the difference between the predicted and actual values.

3. Results and discussion

3.1 Influential factors

The performance of the model in predicting accidents and near misses from the input features is remarkable: the accuracy and precision of the model's predictions are evaluated on the basis of RMSE, which represents the deviation of the residuals and of the correlation coefficient R^2, calculated as follows:

$$RMSE = \sqrt{\frac{\sum(y_{test} - y_{pred})^2}{N-P}} = 0.1 \tag{1}$$

Where: y_{test} = *actual value for the* i_{th} *observation*
y_{pred} = *predicted value for the* i_{th} *observation*
N = number of observations
P = number of parameter estimates

$$R^2 = \frac{Variance\ explained\ by\ the\ model}{Total\ variance} = 0.78 \tag{2}$$

One of the key advantages of the Random Forest Regressor relies in its ability to measure the importance of different features in predicting the target variable. This analysis helps identify the factors and subfactors that have the most significant impact on the occurrence of accidents and near misses. The results of this analysis provide a clear indication of what are the most important elements to be monitored to effectively manage the risks of energy transition. The feature importance is calculated based on the total reduction of the mean squared error (MSE) caused by each feature across all decision trees in the random forest. The importance values are normalized so that the sum of all feature importance is equal to one. For each decision tree, the model calculates a nodes importance using *Gini* Importance. Gini Importance or Mean Decrease in Impurity – MDI - calculates each feature importance as the sum over the number of splits (across all trees) that include the feature, proportionally to the number of samples it splits (in equation 3 an example for a binary tree):

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{3}$$

Where: *nij*= importance of node j
wj = weighted number of samples reaching node j
Cj = impurity value of node j
left(j) = child node from left split on node j
right(j) = child node from right split on node j

The importance for each feature on a decision tree is subsequently calculated as follows:

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k\ in\ all\ nodes} fi_j} \tag{4}$$

Where: *fii* = importance of feature i
nij = importance of node j

Feature importance values are then normalized by dividing them by the sum of all feature importance values. The final feature importance, at the Random Forest level, is the calculated average over all the trees. The sum of the feature's importance value on each tree is calculated and divided by the total number of trees (Eq. 5):

$$RFfi_i = \frac{\sum_{j\ in\ all\ trees} normfi_{ij}}{T} \tag{5}$$

Where: *RFfii* = importance of feature i calculated from all trees in the Random Forest model
normfiij = normalized feature importance for i in tree j
T = total number of trees.

3.2 Detailed feature analysis

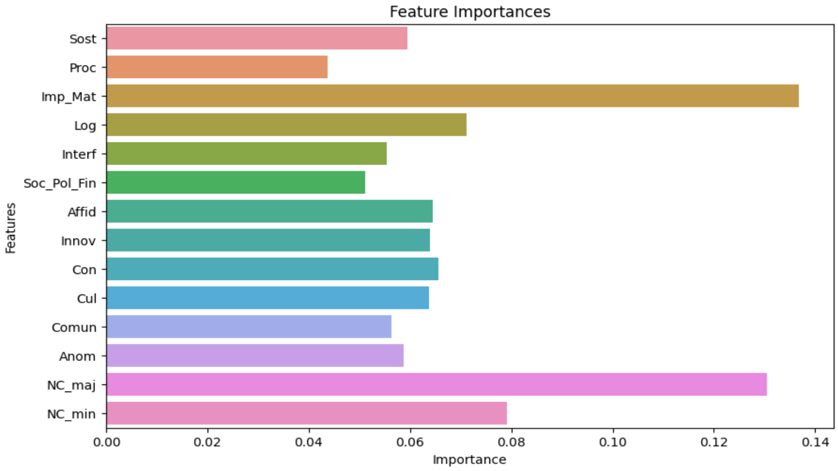

Figure 1: Calculated feature importance.

The feature importance, graphically depicted in Figure 1, allowed revealing the following insights.

Accelerating factors

- Among *Per* (the hazard level), *Imp_Mat* was found to be the most important feature, contributing significantly to the prediction of accidents and near misses. This observation suggests that the technological characteristics of the plants involved in the processes play a crucial role in determining the safety outcomes.
- Among *Con* (the contextual elements), *Log* emerged as a highly influential feature. Factors related to logistical operations and management have a relevant impact on accident and near-miss occurrences.

Mitigating factors

- Among the *Tec* (technical aspects), both *Affid* and *Innov* were found to have notable importance. This result suggests that both the reliability of technical systems and the level of innovation implemented in the processes contribute significantly to safety outcomes.
- Within *Um_org* (human and organizational aspects), *Con* was identified as the most important feature. This finding indicates that the level of knowledge and expertise possessed by individuals and organizations has a strong influence on accident and near-miss occurrences, confirming previous studies in the process industry context relying on different statistical approaches (e.g., Fabiano et al., 2022).

Operative experience data

The distinctly higher importance of major non-conformities (*NC_maj*) highlights the relevant role of non-compliance issues in anticipating the safety incidents.

Table 1: Calculated feature importance.

Feature	Importance
Sost	0.059
Proc	0.044
Imp_Mat	0.137
Log	0.071
Interf	0.055
Soc_Pol_Fin	0.051
Affid	0.064
Innov	0.064
Con	0.066
Cul	0.064
Comun	0.056
Anom	0.059
NC_maj	0.130
NC_min	0.079

Understanding the relative importance of different factors and subfactors provides valuable insights for prioritizing interventions and allocating resources to improve safety. By focusing on the most influential factors, organizations can effectively target their efforts to mitigate risks and prevent accidents and near misses. The percentual importance calculated for each feature in the given applicative case-study is reported in Table 1. It is noteworthy noting that the feature importance analysis is based on the specific dataset and modeling assumptions used in this study and may largely vary in different contexts or datasets. Consequently it is recommended to conduct further analyses and validation studies to ensure the robustness and generalizability of the presented findings.

4. Conclusions

The present study utilized a Random Forest Regressor to analyze the factors influencing safety accidents and near misses. By examining a comprehensive set of factors and subfactors, the key drivers of safety outcomes in the studied context are identified. The main findings of data analysis are summarized as follows.

- The model exhibits a remarkable predictive ability in terms of accuracy and precision (RMSE=0.1, R^2= 0.78).
- The feature importance analysis revealed the significance of factors such as hazard levels, contextual aspects, technical considerations, human and organizational factors, as well as anomalies and non-compliance issues.
- The most important accelerating factor is found to be *Plant and materials characteristics*. This finding confirms that the relevant role of the technological characteristics and loss prevention preventive and mitigating barriers driven by innovation of the process plants in determining the final safety outcomes.
- The distinctly higher importance of *major non-conformities* highlights the relevant role of non-compliance issues in anticipating the safety incidents.

The predictive model developed in this study provides a valuable tool for forecasting safety incidents and near misses based on the identified factors. This research contributes to the existing body of knowledge on safety management and provides practical insights for organizations aiming to improve their safety performance. By understanding and addressing the key drivers of safety incidents, organizations can create safer environments, and mitigate the potential risks associated with accidents and near-misses connected with the energy transition.

Acknowledgments

The authors gratefully acknowledge funding by INAIL, within the framework of the call BRIC/2021/ID3 (Project DRIVERS- Approccio combinato data-driven ed experience-driven all’analisi del rischio sistemico).

References

Bassani A., Vianello C., Mocellin P., Dell’Angelo A., Spigno G., Fabiano B., Maschio G., Manenti F., 2023, Aprioristic integration of process operations and risk analysis: definition of the weighted F&EI-based concept and application to AG2S technology. Ind. Eng. Chem. Res. 62 (1), 500–510.

Breiman L., 2001, Random Forests, Machine Learning, 45, 5-32.

Chen W., Zou W., Zhong K., Aliyeva A., 2023, Machine learning assessment under the development of green technology innovation: A perspective of energy transition, Renewable Energy, 214, 65-73.

Fabiano B., Pettinato M., Currò F., Reverberi A.P., 2022, A field study on human factor and safety performances in a downstream oil industry, Safety Science, 153, 105795.

Gerstorfer Y., Krieg L., Hahn-Klimroth, M., 2023, A notion of feature importance by decorrelation and detection of trends by Random Forest Regression, arXiv preprint arXiv:2303.01156.

Markowski, A.S., Krasławski, A., Vairo, T., Fabiano, B., 2021. Process safety management quality in industrial corporation for sustainable development. Sustainability 13, 9001.

Pasman H., Sripaul E., Khan F., Fabiano B., 2023, Energy transition technology comes with new process safety challenges and risks, Process Safety and Environmental Protection,177, 765-794.

Pasman H.J., Fabiano B., 2021, The Delft 1974 and 2019 European Loss Prevention Symposia: Highlights and an impression of process safety evolutionary changes from the 1st to the 16th LPS, Process Safety and Environmental Protection 147, 80–91.

Schischke A., Papenfuß P., Brem M., Kurz P., Rathgeber A.W., 2023, Sustainable energy transition and its demand for scarce resources: Insights into the German Energiewende through a new risk assessment framework, Renewable and Sustainable Energy Reviews, 176, 113190.

Shah D., Yu Xue Z., Aamodt T.M., 2022, Label encoding for regression networks, In: Proceedings of the Tenth International Conference on Learning Representations, ICLR Publ., Addis Ababa, Ethiopia, arXiv:2212.01927.

Vairo T., Cademartori D., Clematis D., Carpanese M.P., Fabiano B., 2023, Solid oxide fuel cells for shipping: A machine learning model for early detection of hazardous system deviations, Process Safety and Environmental Protection, 172, 84-194.