SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCSD 6215
Sem:……1…… Session:…2024 / 2025……

SECTION A: Project Information.

Program Name: **Masters of Data Science**

Subject Name: **Project 1 (MCSD 6215)**

Student Name: Haslinda binti Abdul Sahak

Metric Number: MCS241004

Student Email & Phone: haslinda45@graduate.utm.my

Project Title: Machine Learning-Driven Accident Prediction for Enhancing Inherent Safety Design Across the Process Design Life Cycle in the Chemical Industry

Supervisor 1:

Supervisor 2 / Industry Advisor(if any):

SECTION B: Project Proposal

Introduction:

The complex processes involved in hazardous chemicals are one of the oldest industries in which safety has been exercised, as it prevented accidents, and protected employees, assets, and the environment thus creating a safe space. The operational dynamics and complexities found in chemical process operations render them difficult to understand in terms of primary causation and prediction of risks. The classic means for risk assessment have been the deterministic and qualitative methods, which are good at presenting and useful for drawing similar judging conclusions but lack the precision and flexibility to address many emerging aspects of process hazards. One of those aspects is going to be, for instance, the combination of machine learning (ML) into predictive transformation perspectives of accidents and safety enhancement.

Machine learning-driven models transform prior data with sophisticated algorithms to reveal previously hidden patterns and appraise incidents much better than conventional systems. Good at instigating the futuristic safety designs into the processing design life cycle; identified live hazards within a value-for-money basis optimized process safety and predicted reduced frequency and severity of accidents (Aziz et al. 2023; Chen & Zhang 2022).

Adding ML to the internal safety of design aligns with where the industry is heading toward predictive and prescriptive safety management systems. The fact that safety principles might be incorporated into the design and improved stepwise processes has a resilience approach to emerging risks. Recent developments in rare-event prediction and machine-learning applications show great potential in filling the gap in traditional risk management systems (Aziz et al., 2023; IEEE, 2023). This paper elaborates on the proposed implementation of machine learning in accident prediction, thus making chemical process industries safer and more efficient.

Problem Background:

Safety in a chemical process design life cycle is crucial because accidents can devastate the company and cause serious casualties. Safety designs have always relied on reactive changes coupled with safety systems that have to be investment costly and ineffective (Kletz, 1978; Hendershot, 2011). However, despite the principal adoption of key safety design (ISD) adoption, severe accidents are still a problem owing to the inadequate implementation complexity of chemical processes (Gupta & Edwards, 2002).

Currently, technology has been found almost completely lacking in its one application towards accident prevention and prediction. This is particularly glaring for machine learning (Kidam et al., 2016a). Machine learning has excellent and powerful analysis treatments for discovering and predicting patterns in extremely big accident data. Unfortunately, though, this potential is yet to be brought into the process design life cycle.

However, the challenge lies in developing a very robust framework for machine learning-enabled accident prediction, thereby improving the inherent safety design of chemical processes. This should systematically identify possible hazards and predict accidents and thus lead to actionable savings to be fed into early design stages. The aim is of course safety improvement but also reduced costs incurred from late design changes and safety interventions (Lutchman et al., 2014). Solving this problem is important for the continuous advancement of safety practices in the chemical industry to ultimately develop safer, more efficient, and sustainable process designs.

Problem Statement:

Most chemical accidents continue to play havoc with the inherent safety design (ISD) principles simply because they do not have early effective detection of hazards. Machine learning as well as advanced predictive technologies are, however, not incorporated well into current practice for identifying imminent threats in the primary design phase, thereby giving rise to a careless attitude towards safety.

Even though vast amounts of accident data are collected in the chemical industry, this data is underutilized when it comes to predicting and preventing future incidents. The existing data analysis techniques are rarely able to uncover deep insights and patterns that can lead to safer design practices. In order to enhance inherent safety, we need a robust framework that uses machine learning to transform raw accident data into actionable intelligence.

Aim of the Project:

Nevertheless, only a scant handful of safety features are included in the product process lifecycle design, but these can be included later without incurring the potential high costs and disruptions in operations associated with developing such technologies. There seems to be a lack of a machine learning model-driven accident prediction before the process is nearly finished, thus leading most of the intervention after the process design has gone substantially to completion. This is how the present work has developed a cost-effective machine learning approach to hazard identification early enough to avoid costly safety modifications later on, with all the operational disruptions that entails.

Objectives of the Project:

The main aim of this research is to realize accident learning in the chemical processing industry (CPI) through the integration of health and safety aspects within the development and design of processes. Specifically, the research intends to achieve:

1. Identify, classify, and rank some major accident enablers as possible root causes traced through the life cycle of process design.
2. Identify, classify, and rank several safety and health recommendations with reference to points of intervention across the process design life cycle.
3. Evaluate accident database information for inherent safety and health and apply findings to
    a. Research and Development
    b. Preliminary Engineering Phase
    c. Basic Engineering Phase
    d. Detailed Engineering Phase
    e. Procurement, Fabrication, Commissioning and Start-up
    f. Operation/ Plant modification.

This will enable CPI safety through accident data becoming directly applicable to design decisions at all stages of the process design life cycle.

Scopes of the Project:

The focus of this study is to sift through a large amount of past accident data and find relevant indications about subsequent safety as well as health integration at the process design phase in the chemical process industry. The scope includes:

1. Data Collection: At least 500,000 accident investigation reports will be compiled by the research from credible sources which include the US Chemical Safety and Hazard Investigation Board (CSB), the US National Transportation Safety Board (NTSB), US Environmental Protection Agency (EPA), Japan's Science and Technology Failure Knowledge Database (FKD), and EU Major Accident Reporting System (e-MARS) all of which house accidents between the dates of 1990 and 2024.

2. Accident Analysis: This study will focus on accidents that are related to process safety so that key contributors can be identified in the chemical process industry. These would be classified according to the following design error category: process condition, reactivity/incompatibility, unsuitable equipment/part, protection, construction material, layout, utility setup, sizing, automation/instrument issues but all would be linked to equipment failure types.

3. Classification of Accident Contributors and Root Causes: Their identification will fall into design, technical, human, organizational, and external errors. Thus, root causes will be classified by the process design life cycle, i.e., research and development (R&D), preliminary engineering, basic engineering, detailed engineering, procurement, fabrication, commissioning start-up, and operation/plant modification.

4. Recommendations Assessment: Such recommendations in the documents shall be broadly classified as preventive, protective, mitigative, and corrective measures. Safety measures can be further classified in terms of inherent safety, inherent health, passive-engineered, and active-engineered measures. Inherent safety and health recommendations will focus on ISD (Inherently Safer Design) principles: minimization, substitution, moderation, and simplification.

5. Statistical Ranking: Statistical analysis will classify and rank the accident contributors and safety recommendations depending on the frequency of occurrence, thus allowing the establishment of critical design-related issues for CPI accidents.

It further addresses the objective of using accident data to facilitate learning, the definition of safety actions, and the implicating integration of such analysis towards inherent safety and health improvements in every phase of process design.

Expected Contribution of the Project:

The research would be of great importance in chemical engineering and process safety in the following respects:

1. Further Development of Safety Prediction Techniques: The research will evolve and validate a machine-learning-based model for predicting accidents in the chemical industry. This model, trained on historical accident data, will generate early warning insights that need to be incorporated into the design process. In turn, predictive methods would change the paradigm from reactive to proactive safety management and, consequently, cut down on both the incidence and severity of accidents associated with the industry (Kidam et al., 2016).

2. Integration of Safety and Health Considerations: The results of the investigation would enable the integration of safety and health considerations systematically in all stages of the life cycle of process design. Making it an integral part of design phases early would build up the way toward the development of more inherently safer and healthier chemical processes. This would improve safety performance and operational effectiveness in total (Hendershot, 2011).

3. Enhanced utilization of Accident Data: The research aims to enhance the use of accident data by transforming it into information with an actionable quality through machine-based learning techniques. It shall elucidate the main accident contributors in the process and recommend safety priorities which will result in a better-informed decision-making process. The advanced use of data will also bring about further improvement in safety practices and compliance with regulatory standards (Lutchman et al, 2014).

4. Advantages for the Economy and Operations: The study will anticipate and prevent hazards later in the design process. It reduces expenditures through preventive measures and avoidance of design changes or safety interventions late in the process. Safety measures can be integrated proactively instead of reactively, increasing the reliability of processes while thereby minimizing disruptions and the overall economic attractiveness of chemical plants (Reniers & Amyotte, 2012).

5. Future Works Benefit: The present study would pave the way for future research on applications of machine learning in process safety and design for inherent safety along with providing insights and advancing further viable methods applicable in other domains with similar concerns. This research should also further the broadening of knowledge on the area of relevance in the intersection of machine learning and industrial safety (Kontogiannis et al, 2016).

Project Requirements:

Software:	• Python or R for data analysis and machine learning model development. • MATLAB for process simulation and advanced mathematical modeling. • Tableau or Power BI for data visualization and reporting.
Hardware:	• High-performance computing (HPC) system or cloud services for training machine learning models. • Workstations with advanced GPUs for model testing and validation.
Technology/Technique/ Methodology/Algorithm:	• **Data Science Techniques**: Data preprocessing, feature engineering, and exploratory data analysis. • **Machine Learning Algorithms**: Regression models, decision trees, random forests, gradient boosting, neural networks. • **Methodology**: Cross-validation, hyperparameter tuning, and model evaluation (accuracy, recall, F1-score).

Type of Project (Focusing on Data Science):

[ ] Data Preparation and Modeling

[ ] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[√] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[√] New

[ ] Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

[√] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: Haslinda binti Abdul Sahak

Signature

27/11/2024

Date

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature **Date**

Name of Supervisor 2 (if any):

Signature **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL

[ ] CONDITIONAL APPROVAL (Minor)

[ ] CONDITIONAL APPROVAL (Major)*

[ ] FAIL*

***** Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

Signature

Date

Name of Evaluator 2:

Signature

Date

Haslinda binti Abdul Sahak.pdf

by Haslinda Binti Abdul Sahak MCS241004

Machine Learning-Driven Accident Prediction for Enhancing Inherent Safety Design Across the Process Design Life Cycle in the Chemical Industry

Introduction:

The complex processes involved in hazardous chemicals are one of the oldest industries in which safety has been exercised, as it prevented accidents, and protected employees, assets, and the environment thus creating a safe workplace. The operational dynamics and complexities found in chemical process operations render them difficult to understand in terms of primary causation and prediction of risks. The classic means for risk assessment have been the deterministic and qualitative methods, which are good at presenting and useful for drawing similar judging conclusions but lack the precision and flexibility to address many emerging aspects of process hazards. One of those aspects is going to be, for instance, the combination of machine learning (ML) into predictive transformation perspectives of accidents and safety enhancement.

Machine learning-driven models transform prior data with sophisticated algorithms to reveal previously hidden patterns and appraise incidents much better than conventional systems. Good at instigating the futuristic safety designs into the processing design life cycle; identified live hazards within a value-for-money basis optimized process safety and predicted reduced frequency and severity of accidents (Aziz et al. 2023; Chen & Zhang 2022).

Adding ML to the internal safety of design aligns with where the industry is heading toward predictive and prescriptive safety management systems. The fact that safety principles might be incorporated into the design and improved stepwise processes has a resilience approach to emerging risks. Recent developments in rare-event prediction and machine-learning applications show great potential in filling the gap in traditional risk management systems (Aziz et al., 2023; IEEE, 2023). This paper elaborates on the proposed implementation of machine learning in accident prediction, thus making chemical process industries safer and more efficient.

Problem Background:

Safety in a chemical process design life cycle is crucial because accidents can devastate the company and cause serious casualties. Inherent safety designs have always relied on reactive changes coupled with safety systems that have to be investment costly and ineffective (Kletz, 1978; Hendershot, 2011). However, despite the principal adoption of key safety design (ISD) adoption, severe accidents are still a problem owing to the inadequate implementation complexity of chemical processes (Gupta & Edwards, 2002).

Currently, technology has been found almost completely lacking in its one application towards accident prevention and prediction. This is particularly glaring for machine learning (Kidam et al., 2016a). Machine learning has excellent and powerful analysis treatments for discovering and predicting patterns in extremely big accident data. Unfortunately, though, this potential is yet to be brought into the process design life cycle.

However, the challenge lies in developing a very robust framework for machine learning-enabled accident prediction, thereby improving the inherent safety design of chemical processes. This should systematically identify possible hazards and predict accidents and thus lead to actionable savings to be fed into early design stages. The aim is of course safety improvement but also reduced costs incurred from late design changes and safety interventions (Lutchman et al., 2014). Solving this problem is important for the continuous advancement of safety practices in the chemical industry to ultimately develop safer, more efficient, and sustainable process designs.

Problem Statement:

Most chemical accidents continue to play havoc with the inherent safety design (ISD) principles simply because they do not have early effective detection of hazards. Machine learning as well as advanced predictive technologies are, however, not incorporated well into current practice for identifying imminent threats in the primary design phase, thereby giving rise to a careless attitude towards safety.

Even though vast amounts of accident data are collected in the chemical industry, this data is underutilized when it comes to predicting and preventing future incidents. The existing data analysis techniques are rarely able to uncover deep insights and patterns that can lead to safer design practices. In order to enhance inherent safety, we need a robust framework that uses machine learning to transform raw accident data into actionable intelligence.

Aim of the Project:

Nevertheless, only a scant handful of safety features are included in the product process lifecycle design, but these can be included later without incurring the potential high costs and disruptions in operations associated with developing such technologies. There seems to be a lack of a machine learning model-driven accident prediction before the process is nearly finished, thus leading most of the intervention after the process design has gone substantially to completion. This is how the present work has developed a cost-effective machine learning approach to hazard identification early enough to avoid costly safety modifications later on, with all the operational disruptions that entails.

Objectives of the Project:

The main aim of this research is to realize accident learning in the chemical processing industry (CPI) through the integration of health and safety aspects within the development and design of processes. Specifically, the research intends to achieve:

1. Identify, classify, and rank some major accident enablers as possible root causes traced through the life cycle of process design.
2. Identify, classify, and rank several safety and health recommendations with reference to points of intervention across the process design life cycle.
3. Evaluate accident database information for inherent safety and health and apply findings to
    a. Research and Development
    b. Preliminary Engineering Phase
    c. Basic Engineering Phase
    d. Detailed Engineering Phase
    e. Procurement, Fabrication, Commissioning and Start-up
    f. Operation/ Plant modification.

This will enable CPI safety through accident data becoming directly applicable to design decisions at all stages of the process design life cycle.

Scopes of the Project:

The focus of this study is to sift through a large amount of past accident data and find relevant indications about subsequent safety as well as health integration at the process design phase in the chemical process industry. The scope includes:

1. Data Collection: At least 500,000 accident investigation reports will be compiled by the research from credible sources which include the US Chemical Safety and Hazard Investigation Board (CSB), the US National Transportation Safety Board (NTSB), US Environmental Protection Agency (EPA), Japan's Science and Technology Failure Knowledge Database (FKD), and EU Major Accident Reporting System (e-MARS) all of which house accidents between the dates of 1990 and 2024.

2. Accident Analysis: This study will focus on accidents that are related to process safety so that key contributors can be identified in the chemical process industry. These would be classified according to the following design error category: process condition, reactivity/incompatibility, unsuitable equipment/part, protection, construction material, layout, utility setup, sizing, automation/instrument issues but all would be linked to equipment failure types.

3. Classification of Accident Contributors and Root Causes: Their identification will fall into design, technical, human, organizational, and external errors. Thus, root causes will be classified by the process design life cycle, i.e., research and development (R&D), preliminary engineering, basic engineering, detailed engineering, procurement, fabrication, commissioning start-up, and operation/plant modification.

4. Recommendations Assessment: Such recommendations in the documents shall be broadly classified as preventive, protective, mitigative, and corrective measures. Safety measures can be further classified in terms of inherent safety, inherent health,

passive-engineered, and active-engineered measures. Inherent safety and health recommendations will focus on ISD (Inherently Safer Design) principles: minimization, substitution, moderation, and simplification.

5. Statistical Ranking: Statistical analysis will classify and rank the accident contributors and safety recommendations depending on the frequency of occurrence, thus allowing the establishment of critical design-related issues for CPI accidents.

It further addresses the objective of using accident data to facilitate learning, the definition of safety actions, and the implicating integration of such analysis towards inherent safety and health improvements in every phase of process design.

Expected Contribution of the Project:

The research would be of great importance in chemical engineering and process safety in the following respects:

1. Further Development of Safety Prediction Techniques: The research will evolve and validate a machine-learning-based model for predicting accidents in the chemical industry. This model, trained on historical accident data, will generate early warning insights that need to be incorporated into the design process. In turn, predictive methods would change the paradigm from reactive to proactive safety management and, consequently, cut down on both the incidence and severity of accidents associated with the industry (Kidam et al., 2016).

2. Integration of Safety and Health Considerations: The results of the investigation would enable the integration of safety and health considerations systematically in all stages of the life cycle of process design. Making it an integral part of design phases early would build up the way toward the development of more inherently safer and healthier chemical processes. This would improve safety performance and operational effectiveness in total (Hendershot, 2011).

3. Enhanced utilization of Accident Data: The research aims to enhance the use of accident data by transforming it into information with an actionable quality through machine-based learning techniques. It shall elucidate the main accident contributors in the process and recommend safety priorities which will result in a better-informed decision-making process. The advanced use of data will also bring about further improvement in safety practices and compliance with regulatory standards (Lutchman et al, 2014).

4. Advantages for the Economy and Operations: The study will anticipate and prevent hazards later in the design process. It reduces expenditures through preventive measures and avoidance of design changes or safety interventions late in the process. Safety measures can be integrated proactively instead of reactively, increasing the reliability of processes while thereby minimizing disruptions and the overall economic attractiveness of chemical plants (Reniers & Amyotte, 2012).

5. Future Works Benefit: The present study would pave the way for future research on applications of machine learning in process safety and design for inherent safety along with providing insights and advancing further viable methods applicable in other domains with similar concerns. This research should also further the broadening of knowledge on the area of relevance in the intersection of machine learning and industrial safety (Kontogiannis et al, 2016).

Haslinda binti Abdul Sahak.pdf

ORIGINALITY REPORT

2%	1%	1%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universiti Teknologi Malaysia Student Paper	1%
2	www.researchgate.net Internet Source	1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		