2024 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET), Faculty of Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh, 07-08 December, 2024

ML Based Ammonia Converter Leakage Detector and Trip Predictor For Improving Safety

Mritunjoy Sarker
Process Safety Officer
Libyan Fertilizer Company, Libya
Sarkermritunjoy89@gmail.com

Md. Mahbubur Rahman
Electrical and Electronic Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh 2224, Bangladesh
mahbubur@jkkniu.edu.bd

MD. Shameem Ahammed
Electrical and Electronic Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh 2224, Bangladesh
shameem_19102913@jkkniu.edu.bd

Sourav Sarker
Electrical and Electronic Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh 2224, Bangladesh
souravsarker50.eee@gmail.com

Zinia Jahan
Electrical and Electronic Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh 2224, Bangladesh
ziniajahan2941@gmail.com

Md. Shake Farid Uddin
Assistant Manager(Software Engineer)
Nuclear Power Plant Company
Bangladesh Limited.
Dhaka, Bangladesh
sfuddin.iit@gmail.com

2024 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET) | 979-8-3315-0947-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICRPSET64863.2024.10955943

***Abstract*— An Ammonia converter is a crucial and sophisticated vessel in an ammonia plant, where Ammonia gas produced by reacting with Hydrogen and Nitrogen. It is highly recommended to ensure maximum safety for continuous process without shutdown the ammonia converter. Machine learning based safety system enhances the plant safety, minimizes downtime, and reduces operational costs in an ammonia plant. We proposed a machine learning model with Naive Bayes algorithm which is a supervised machine learning. To train the model, we use real time operational data collected from an ammonia plant. The data will be processed and analyzed to train the model with train-test ratio of 80:20. We identify the best parameter for the algorithm using GridSearchCV and five-fold cross validation and tune the hyperparameter for this model. The highest accuracy score of 93.33% gained by this model. The model detect leakage and early predict abnormal condition of the ammonia converter which improved the safety of the Ammonia plant.**

***Keywords*— Machine Learning, Naive Bayes, Ammonia Converter, Abnormal condition detector..**

I. INTRODUCTION

Ammonia is one of the most widely produced chemicals in the world, with a significant portion dedicated to the production of fertilizers, a cornerstone of modern agriculture[1]. Central to the ammonia production process is the ammonia converter, a high-pressure, high-temperature vessel where nitrogen and hydrogen react to form ammonia gas[2][3]. Given the extreme operating conditions and the critical nature of the converter, ensuring its safe and efficient operation is of paramount importance. Even minor malfunctions, such as leaks or abnormal temperature variations, can pose significant safety risks, lead to unplanned effective hyperparameters for accurate predictions. shutdowns, and result in costly downtime and maintenance[4].

In industrial environments like ammonia plants, maintaining operational continuity while ensuring safety is a persistent challenge. Traditional safety mechanisms, although reliable, often react to incidents rather than predict and prevent them[5][6]. Recent advancements in data-driven technologies, particularly machine learning (ML), have paved the way for more proactive safety systems. In ammonia plants, ML-driven predictive models can be applied to monitor critical parameters like temperature, pressure, gas concentrations, and equipment status. For example, a model can learn from historical incidents where pyrophoric materials, such as certain catalysts, became unstable. By analyzing vast amounts of real-time operational data, machine learning models can detect subtle patterns indicative of potential failures, enabling predictive maintenance and early interventions[7][8][9].

This research focuses on the application of a machine learning model using the Naive Bayes algorithm to improve safety in ammonia plants. The Naive Bayes algorithm, a popular supervised learning method, is particularly suited for industrial safety applications due to its simplicity, robustness, and ability to handle large datasets with varying conditions. Using real-time data from an ammonia plant, the proposed model is trained to detect abnormal conditions and predict potential leaks in the ammonia converter. The model's performance is optimized through GridSearchCV and five-fold cross-validation, ensuring the selection of the most effective hyperparameters for accurate predictions..

The results of this study are promising, with the model achieving an accuracy rate of 93.33%, demonstrating its potential to significantly enhance plant safety. By predicting abnormal conditions early and minimizing the risk of leaks, this model not only contributes to improved operational safety but also reduces downtime and operational costs, ensuring a more efficient and secure ammonia production process.

The rest of the paper are organized as follows: Section II provides a detailed explanation of the proposed model, including data collection, data pre-processing, exploratory data analysis, dataset splitting, defining param grid, and hyperparameter tuning. Section III discuss about the result of proposed model. Section IV conclude the paper with some future works.

II. PROPOSAL MODEL

In this section, the Naïve Bayes method is presented for detecting abnormal condition in an Ammonia converter. The model designed by collecting datasets and using them to train our model to determine whether the converter in risk-free or not more accurately. The Figure 3 shows the working procedure of the proposed method.

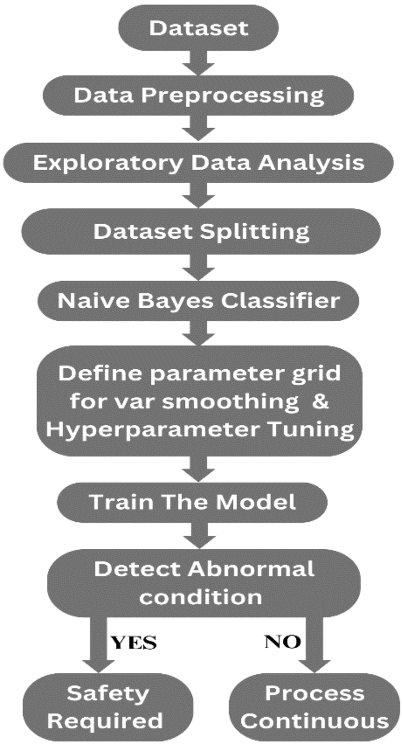

Figure 3: Working procedure of the proposed method

A. Dataset

The dataset is collected from the UHDE process based 1200MT/D Ammonia Plant-1 of Libyan Fertilizer Company (LIFECO, Libya). It has 300 instances, including 12 features and 1 targeted value. There are two types of targeted value. One is 0, which means normal condition of Ammonia converter, and another is 1 means abnormal condition. It contains information about sensor data in real time of an Ammonia convert**.** The features are ***Pressure*** (PI-57017, Ammonia Converter Inlet and Outlet) in kg/cm2, ***Differential Pressure*** (PDI-57015, Difference between Converter Inlet and Outlet **),** ***1st bed temperature*** (TR-57051, Converter first bed outlet), ***2nd bed temperature*** (TR-57054, Converter second bed outlet), ***3rd bed temperature*** (TR-57056, Converter third bed outlet), ***Recycle Flow*** (FR-57001, converter inlet gas flow) in nm3/hr, ***Make-up Flow*** (FR-57006, converter inlet gas flow) in nm3/hr, ***Syngas Comp RPM*** (SI-56001, speed of syngas comp), ***% Of H2*** (AR-57001, Hydrogen Analyzer), ***% Of N2*** (AR-57002, Nitrogen Analyzer), ***% Of CH4*** (AR-57008, Methen Analyzer), ***% Of NH3*** (AR-57003, Hydrogen Analyzer), ***% Of Ar*** (AR-57004 , Argon Analyzer) and the target value which means the converter in normal condition or not.

The portion of our dataset are shown in Figure 2.

	Inlet Pressure (kg/cm2)	Outlet Pressure (kg/cm2)	Differential Pressure	T1	T2	T3	Recycle Flow(nm3/hr)	Make-up Flow(nm3/hr)	Syngas Comp RPM	% Of H2	% of N2	% of CH4	% of Ammonia	% of Ar	Target
518	205.49	200	5.50	518	466	467	370000.0	125000.0	11958	61.92	23.56	7.52	3.41	3.60	1
519	199.99	195	5.00	518	464	465	365000.0	112000.0	11801	61.95	23.87	7.35	3.56	3.24	0
520	205.70	200	5.68	519	466	467	370000.0	125000.0	11958	61.31	24.24	7.51	3.40	3.56	1
521	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1

Figure 2: Dataset

The distribution of target value is shown in Figure 1.

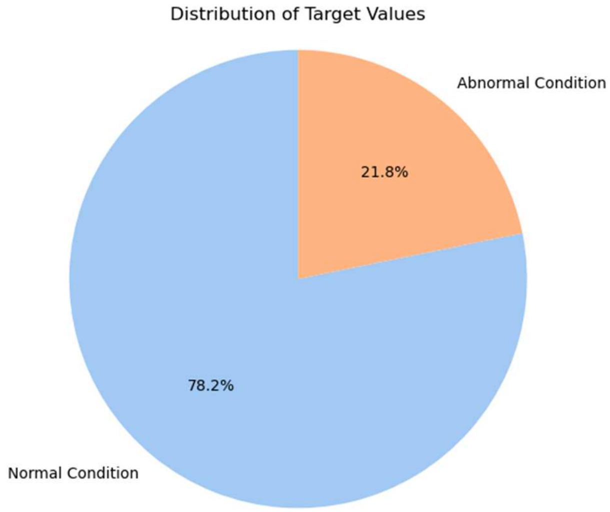

Figure 1: Distribution of Target value

Data pre-processing and analysis are commonly used in Python libraries. We imported Pandas (To analyze data), Numpy (used for working with array), matplotlib (serves static, interactive visualization), and Seaborn (to visualize random distributions) libraries.

B. Data Preprocessing

Here, we checked whether our dataset has missing values, duplicate values, or noisy data and found two duplicate values whose are eliminated and no missing value.

C. Exploratory Data Analysis

In this section, we have seen that our dataset has 523 data where 409 have in normal condition and the remaining are in faulty condition. To understand the targeted values clearly Figure 4 shows the correlation heatmap of each feature with the others.

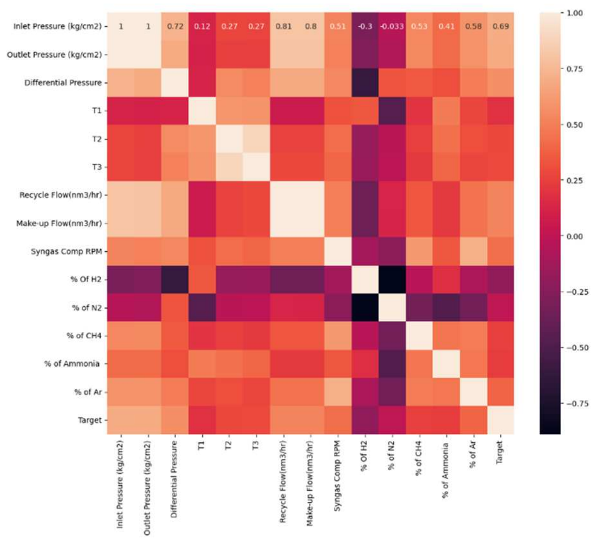

Figure 4: Correlation Heatmap

To visualize the relationship of every feature with each other in the dataset the pair plot is generated in Figure 5. It combines scatter plots and histograms at a time and also gives an outline of correlations and distributions of the datasets. Here diagonal plots are known as histograms which show the distribution of a single feature. The Off-diagonal plots known as scatter plots show the relationship between two features0.

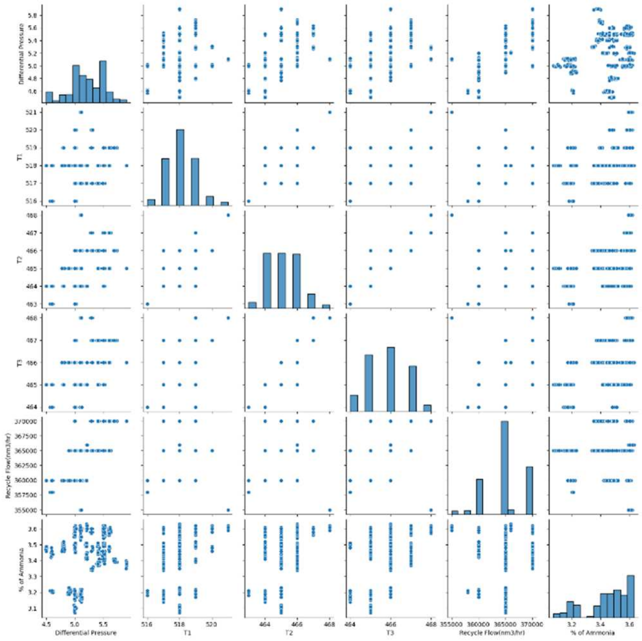

Figure 5: Pair Plot

D. Splitting Dataset

In this step, we split our dataset into two parts, i.e. training and testing data. The ratio of train and test data is 80:20. So. to train our model we use 239 data and for testing the sharpness of our system we use 59 data.

E. Naïve Bayes Classifier

A probabilistic machine learning algorithm based on Bayes' Theorem, Naive Bayes, enables data to be classified based on the likelihood of an event occurring. Specifically, it can be used to categorize input data into different classes based on historical data and probability estimations. It is possible to predict the probability of a leak or trip event by using Naive Bayes based on sensor readings and environmental parameters at the ammonia plant.

Bayes' Theorem is mathematically expressed as:

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$$

where,

- P(A|B) is the **posterior probability**: the probability of occurrence A (e.g., leakage or trip) given incidents B (e.g., sensor readings).
- P(B|A) is the **likelihood**: the probability of the facts B given that A is true.
- P(A) is the **prior probability**: the probability of event A occurring.
- P(B) is the **marginal likelihood**: the probability of the evidence B under all possible hypotheses.

According to the Naive Bayes algorithm, every variable—such as sensor readings—is conditionally independent, which means that each feature affects the result (leak or no leak) separately. Although real-world data frequently contradicts this hypothesis, Naive Bayes typically performs well in this operations.

1) Define the parameter grid for var_smoothing: In order to identify which value provides the optimum trade-off between model performance, variance management, and generalization to new data, the parameter grid for var_smoothing allows in the investigation of a range of values. Building an efficient, trustworthy machine learning model requires fine-tuning var_smoothing as it is important in controlling the model's response to data variability.
We have found the best parameter grid for var_smoothing 'var_smoothing': 0.012328467394420659.

2) Hyperparameter Tuning: GridSearchCV is used to pinpoint the perfect combination of hyperparameters that can increase our machine learning model's performance.

Cross-validation is used to partition the train data into two parts i.e. train and validation data. 5-fold (K=5) cross validation which divides the train data into five(5) partitions used by us. Which uses four(4) partitions for training and one(1) partition for testing in each iteration and continues 5 times. In the end, it gives the average performance0.

For hyperparameter tuning, we use GridSearchCV with 5-fold cross-validation. The arguments are as follows:

- estimator is A scikit-learn model which trained
- param grid is a list of parameter values
- scoring is the performance measure
- cv is the number of folds for k-fold cross-validation
- verbose controls the verbosity; the higher, the more messages.

Now, We train the Naïve Bayes classifiers with the above arguments of GridSearchCV with 5-fold cross-validation and got an accuracy of 93.33%.

Confusion Matrix sum up the achievement of a machine learning classifier basis of test data. It shows the number of sampling. Here the matrix format in Table 1.

True Positive(TP)	False Negative(FN)
False Positive(FP)	True Negative(TN)

Table 1: CONFUSION MATRIX

Where,

- TP refers to truly predict positive outcomes.
- TN refers to truly predict negative outcomes.
- FP stands for falsely predict positive outcomes
- FN stands for falsely predict negative outcomes.

Logistic Regression classifier performs well when the data is linearly separate. It is interpretable; we can determine how each input information influences predictions by examining feature coefficients. The confusion matrix (Figure 6) of Logistic Regression classifier for the dataset is given below:

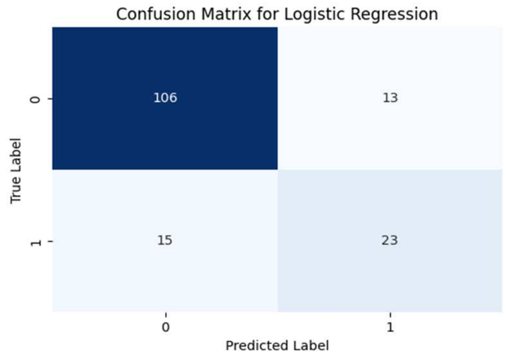

Figure 6: Confusion Matrix of Logistic Regression classifier

Support Vector Machine (SVM) classifier is appropriate in situations when we have few ammonia leakage records since it is reliable with smaller datasets. The confusion matrix (Figure 7) of Logistic Regression classifier for the dataset is given below:

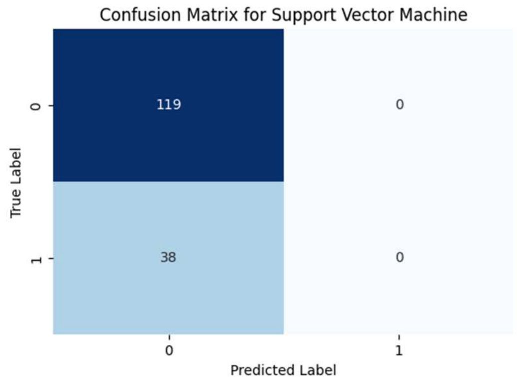

Figure 7: Confusion Matrix of Support Vector Machine classifier

The confusion matrix (Figure 8) of our proposed model for the dataset is given below.

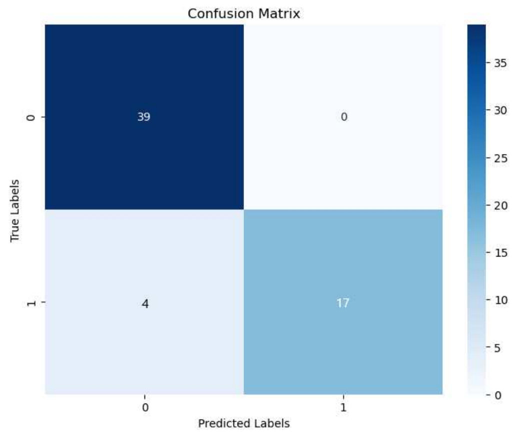

Figure 8: Confusion Matrix of Naïve Bayes classifier

Naive Bayes is computationally simple and quick, making it suitable for real-time monitoring and prediction. In an ammonia plant, quick decisions can be crucial to avoiding safety incidents. It is effective with small to medium-sized datasets, making it ideal if there's limited historical data on leak occurance. It can handle both continuous sensor readings (pressure or temperature) and categorical data (system status: "operating" or "shutdown"). The model provides probabilities, which can be used to assess risk levels. This is useful for decision-making in safety systems, as it can help plant operators gauge the severity of a potential issue.

For performance comparison, we illustrate a bar chart on Figure 9. Here are shown all the model performance that we compared.

Figure 9: Performance Comparison

III. RESULT AND DISCUSSION

In this part, we will discuss the result of our implemented model and compare it with other four machine learning techniques. A ROC (Receiver Operating Characteristic) curve is a graph that shows the achievement of a model at all classification thresholds**Error! Reference source not found.**. It uses two parameters:

- True Positive Rate(TPR):

$$\text{TPR} = \frac{TP}{TP + FN}$$

- False Positive Rate(FPR):

$$\text{FPR} = \frac{FP}{FP + TN}$$

Area Under the Curve denotes as AUC. To compute the points in an ROC curve of different classification thresholds provided by AUC. It counts the entire 2D area underneath the ROC 0. The ROC curve (Figure 10) with AOC values for our proposed model is given below.

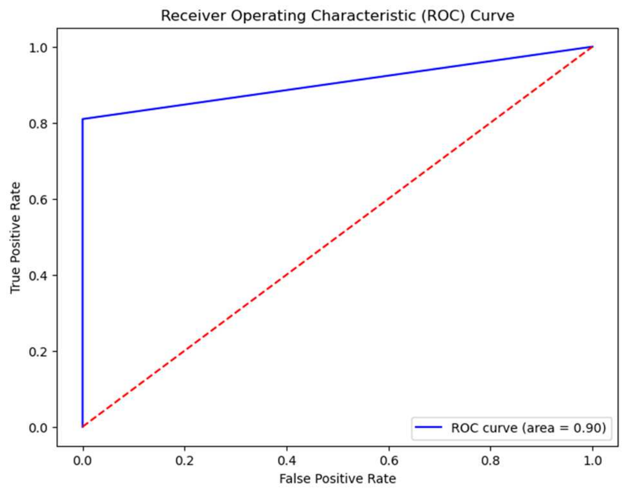

Figure 10: ROC curve with AUC values

Accuracy means how close the predicted and actual values are.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision measures true positive prediction among all positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the positive prediction among all actual positive items.

$$Recall = \frac{TP}{TP + FN}$$

f1-score is a metric that balances precision and recall.

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Now, the classification report for training and test data of proposed model including precision, recall, f1-score, and accuracy are given below.

Target	Precision (%)	Recall (%)	f1-score (%)	Accuracy (%)
0	99	100	99	99.16
1	100	95	98	

Table 2: Classification Report for Training Data

Table 2 shows the accuracy of training data for the Naïve Bayes classifier which gained 99.16%.

Target	Precision (%)	Recall (%)	f1-score (%)	Accuracy (%)
0	91	100	95	93.33
1	100	81	89	

Table 3: Classification Report for Test Data

Table 3 shows the accuracy of test data for the Naïve Bayes method which obtained 93.33%.

A Naive Bayes machine learning model was trained on operational data from an ammonia converter, using an 80:20 split for training and testing to ensure effective learning and evaluation. Hyperparameter tuning was performed with GridSearchCV and five-fold cross-validation, optimizing the model for a high accuracy of 93.33% in detecting leaks and abnormal conditions. The model's performance was assessed through precision, recall, and F1 score metrics.

Now, the accuracy of all models including precision, recall, f1-score, and AUC values are given below (Table 4).

ML Model	Accuracy (%)	Precision (%)	Recall (%)	f1-score (%)	AUC
Naïve Bayes	**93.33**	97.50	94.00	95.25	0.90
Logistic Regression	82.16	87.60	89.07	88.33	0.93
Support Vector Machine	75.80	75.80	1.00	86.23	0.92

Table 4: Result Analysis

Table 4 shows the detail experimental results among different machine learning method. It shows that the Naïve Bayes method obtained the maximum accuracy of 93.33% and AUC values of 0.90. Also, it appears the logistic regression, Support Vector Machine classifiers have accuracy of 82.16%, 75.80% and AUC values are 0.93, 0.92 respectively. But Logistic Regression method is vulnerable to overfitting in circumstances that there are too many variables and not adequate information. Also Support Vector Machine is computationally complicated, particularly when dealing with big datasets, less interpretable since it is difficult to examine kernel adjustments. Although, Naïve Bayes Model , demonstrating its effectiveness in enhancing plant safety, reducing false alarms, and improving operational efficiency.

IV. Conclusion

To enhance safety in an ammonia plant, we developed a machine learning model using the Naive Bayes algorithm, which was applied to the ammonia converter. As a result of utilizing real-time operational data, the model was trained to predict abnormal conditions and detect leaks, which is necessary for preventing potential hazards and ensuring the safe and continuous operation of the plant. In this study, we demonstrate the potential of machine learning in industrial safety applications, as the model contributes to the reduction of accidents, enhances operational continuity, and lowers maintenance costs. In the future, this approach could be extended by integrating additional machine learning. Additionally, exploring alternative machine learning algorithms—including deep learning models like LSTMs for time-series prediction and ensemble methods like Gradient Boosting—could further enhance the model's predictive accuracy and reliability. A more comprehensive safety solution may also be achieved by incorporating the model into other critical systems across the plant in real-time. Overall, this study demonstrates the viability and importance of predictive safety models in industrial settings, especially in high-risk environments such as ammonia plants.

References

[1] D. Saygin *et al.*, “Ammonia Production from Clean Hydrogen and the Implications for Global Natural Gas Demand,” *Sustain. 2023, Vol. 15, Page 1623*, vol. 15, no. 2, p. 1623, Jan. 2023.

[2] L. Jiang, J. Yu, S. Wang, X. Wang, L. Schwark, and G. Zhu, “Complete ammonia oxidization in agricultural soils: High ammonia fertilizer loss but low N2O production,” *Glob. Chang. Biol.*, vol. 29, no. 7, pp. 1984–1997, Apr. 2023.

[3] O. A. Ojelade, S. F. Zaman, and B. J. Ni, “Green ammonia production technologies: A review of practical progress,” *J. Environ. Manage.*, vol. 342, p. 118348, Sep. 2023.

[4] I. Nallathambi *et al.*, “Impact of Fireworks Industry Safety Measures and Prevention Management System on Human Error Mitigation Using a Machine Learning Approach,” *Sensors 2023, Vol. 23, Page 4365*, vol. 23, no. 9, p. 4365, Apr. 2023.

[5] Y. Wang and S. H. Chung, “Artificial intelligence in safety-critical systems: a systematic review,” *Ind. Manag. Data Syst.*, vol. 122, no. 2, pp. 442–470, Feb. 2022.

[6] M. M. Rahman, A. Hossain, and M. A. Sikder, “Machine Learning Applications in Industry Safety: Analysis and Prediction of Industrial Accidents,” *Int. Conf. Smart Syst. Appl. Electr. Sci.*

ICSSES 2024, 2024.

[7] N. Khan and S. A. Ammar Taqvi, “Machine Learning an Intelligent Approach in Process Industries: A Perspective and Overview,” *ChemBioEng Rev.*, vol. 10, no. 2, pp. 195–221, Apr. 2023.

[8] K. P Sunilkumar and S. P A, “Ensuring Safety in Ammonia Production: A Comprehensive Hazard Analysis of the Reforming Section,” *SSRN Electron. J.*, Jul. 2023.

[9] C. Zhang, Z. Wang, X. Chen, and Y. Xiang, “Risk assessment of liquid ammonia tanks based on Bayesian network and Probit model,” *Process Saf. Prog.*, vol. 43, no. 2, pp. 287–298, Jun. 2024, doi: 10.1002/PRS.12560.

[10] Ahluwalia, H. (2024, March 19). Pair plots in machine learning. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2024/02/pair-plots-in-machinelearning/

[11] Shah, R. (2024, July 9). Tune hyperparameters with GridSearchCV. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/06/tune-hyperparameters-with-gridsearchcv.

[12] Google for Developers. (n.d.). Classification: ROC curve and AUC. *Google for Developers*.

https://developers.google.com/machine-learning/crashcourse/classification/roc-and-auc