

Smoke Detection Using Machine Learning algorithms

B SAI MUKESH REDDY

224107002

Abstract:

Smoke detection is very important in chemical industries. The formation of smoke means something is wrong with the system and this has to be taken care immediately otherwise something catastrophic may happen. In this study we train some **classification** algorithms and see how the accuracy for each model changes and the one with highest accuracy is taken for our use case.

Introduction:

SMOKE DETECTION IS CRITICALLY IMPORTANT IN CHEMICAL INDUSTRIES FOR SEVERAL REASONS:

Early warning of potential fires: Chemical industries often handle flammable materials and operate at high temperatures, making them more susceptible to fires. Smoke detectors can provide early warning of a potential fire, allowing for faster response times and a greater chance of extinguishing the fire before it gets out of control.

Protection of people and property: In addition to the risk of fire, chemical industries can also produce toxic or hazardous gases that can be harmful to people and the environment. Smoke detectors can provide early warning of these hazards, allowing for prompt evacuation and minimizing the risk of injury or death.

Compliance with regulations: Chemical industries are subject to a wide range of regulations and standards related to safety, health, and environmental protection. Many of these regulations require the use of smoke detectors as a key element of fire safety systems.

Minimization of downtime and financial losses: Fires and other incidents can cause significant disruptions to operations and lead to financial losses due to property damage, downtime, and potential liability. Smoke detectors can help minimize these losses by providing early warning of potential hazards and allowing for prompt response and mitigation.

Overall, smoke detection is an essential component of fire safety and hazard prevention in chemical industries, helping to protect people, property, and the environment.

MACHINE LEARNING CAN BE USED FOR SMOKE DETECTION IN VARIOUS WAYS, DEPENDING ON THE AVAILABLE DATA AND THE SPECIFIC APPLICATION. HERE ARE SOME COMMON APPROACHES:

Image or video analysis: Machine learning algorithms can be trained on large datasets of images or videos to recognize the visual patterns of smoke. This approach can be applied to smoke detectors that have cameras or other visual sensors. The algorithm can learn to recognize the color, shape, and movement patterns of smoke and distinguish it from other sources of visual

disturbance, such as steam or dust. When the algorithm detects smoke in the image or video data, it can trigger an alarm.

Sensor data analysis: Machine learning can also be used to analyze data from other types of sensors, such as temperature or gas sensors, to detect the presence of smoke. The algorithm can learn to recognize the patterns of sensor data that indicate a potential fire, such as a sudden increase in temperature or the presence of specific gases. This approach can be used to monitor environments where visual smoke detection may be less effective, such as those with high ceilings or strong air flow.

Fusion of multiple data sources: Machine learning can also be used to fuse data from multiple sources, such as visual and sensor data, to improve the accuracy and reliability of smoke detection. The algorithm can learn to integrate the different types of data and recognize the patterns that indicate the presence of smoke or a potential fire. This approach can be particularly useful in complex environments where multiple sources of data are available, and traditional smoke detectors may be less effective.

Overall, machine learning can improve the accuracy and speed of smoke detection, which is critical for ensuring the safety of people and property in a wide range of environments, including chemical plants, industrial facilities, and residential buildings. However, careful training and validation of the machine learning algorithms are necessary to ensure their reliability and effectiveness.

The smoke detection can be done using machine learning classification algorithms like logistic regression knn classifier and decision tree.

Methodology:

The dataset has input features like: UTC, Temperature C, Humidity % , TVOC ppb , eCO2 ppm , Raw H2 , Raw Ethanol , Pressure hPa , PM1.0 , NC2.5 , CNT .

The output is of the column Fire Alarm which has a Boolean yes or no as the output.

Data preprocessing:

- finding the null values and the data set has no null values.
- finding the duplicate values and the data set has no duplicate values.
- finding the mean and standard deviation of every feature
- plotting all the features using scipy stats library
- check for outliers
- plot the correlation matrix (spearman correlation is used)
- feature scaling as the features are of various ranges and to bring uniformity to the data all the features are scaled using the minmaxscaler

the train test split is made for the whole data (train data points: 26763, test data points :8921) and the ratio of this split is 0.25 which is a hyper parameter that can be tuned if needed and random state is set so as to work with the same test data again.

Model 1:

We use a logistic regression algorithm with the hyper parameters penalty function, C and max iter after running this algorithm we get the best parameters by performing the grid cv model function

Best parameters are: c 0.5, max iter 1, penalty l2

the classification report is as follows:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	1.00	0.87	0.93	8921
accuracy			0.87	8921
macro avg	0.50	0.43	0.46	8921
weighted avg	1.00	0.87	0.93	8921

Model 2:

The same can be repeated with another algorithm called knn

Here we have hyper parameters as n neighbors

After using the hyper parameter method available in python like gridsearchcv we get this to be 50 as optimum.

After implementing this model, we get the results as 100% accuracy on test data.

A same model with decision tree is also implemented and has resulted an accuracy of 100%.

Results:

The correlation between features is as follows:

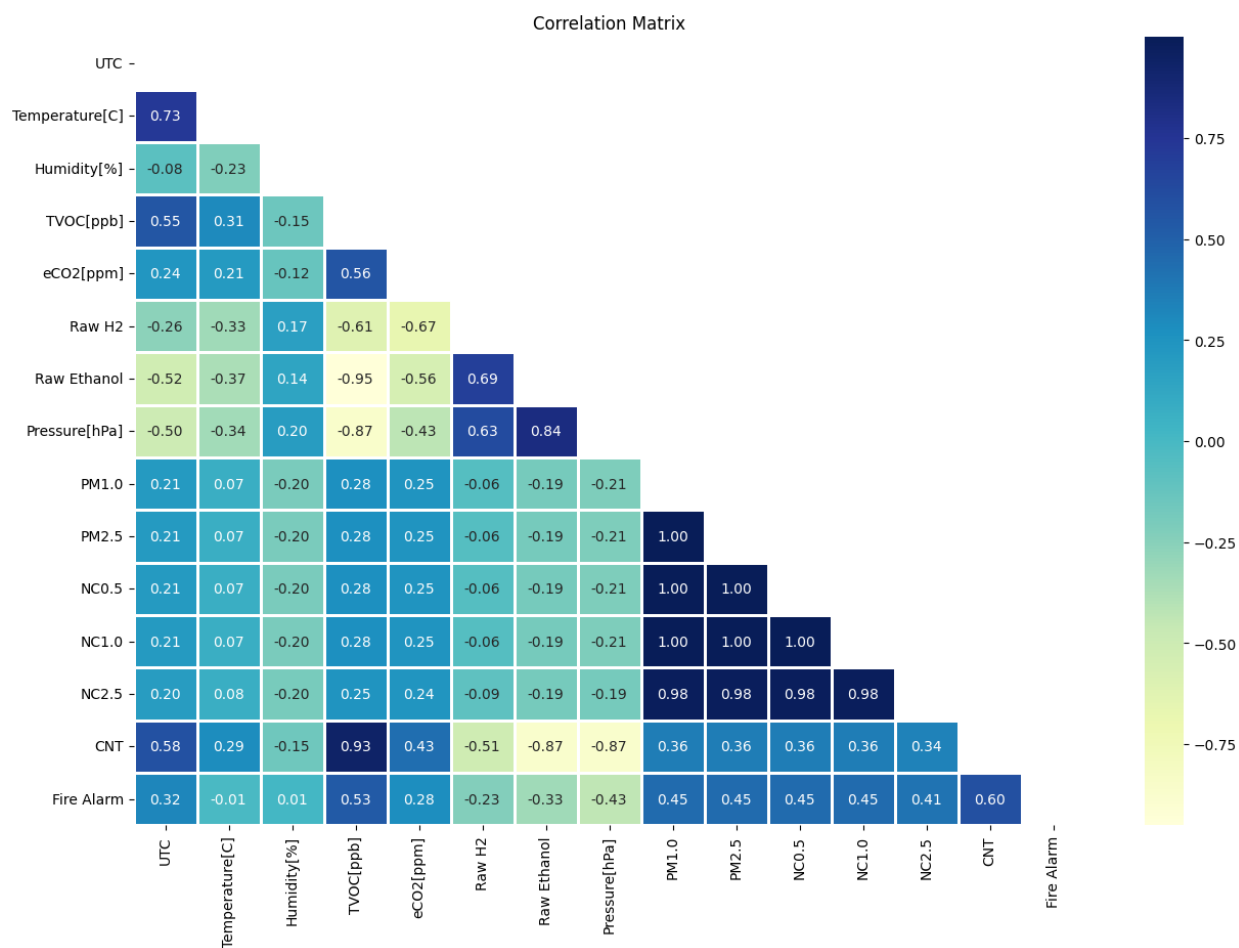


Figure 1Correlation plot

The accuracy values for all the models have been reported as shown in the figure below:

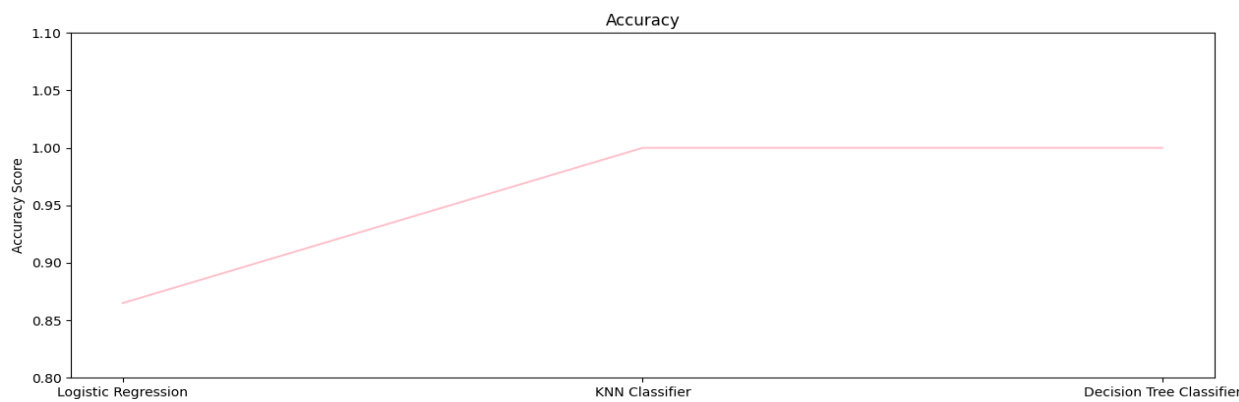


Figure 2 accuracy plot

This model has 100% accuracy and detect the smoke at any point of time as this is needed in real industrial applications as 0.1% errors can also make catastrophic events leading to death of many people and property loss.

Conclusion:

The model helps detect the smoke in industries by detecting concentrations of gas and fire causing chemicals released into air. The data sample size is less and we cannot say that 100% is true for a new test case and for a new test case if this model fails we need to reevaluate the model and process the sensor again as machine learning is a continuous development model. We can also make this model using CNN RNN neural network model if we need more accuracy.

References:

NAMOZOV, A. & CHO, Y.. (2018). An Efficient Deep Learning Algorithm for Fire and Smoke Detection with Limited Data. Advances in Electrical and Computer Engineering. 18. 121-128. 10.4316/AECE.2018.04015.

[DeepSmoke: Deep learning model for smoke detection and segmentation in outdoor environments - ScienceDirect](#)

Appendix:

Jupiter notebook file: [CL653Courseworkproject/smokedetection.ipynb at main · BSAIMUKESHREDD/CL653Courseworkproject · GitHub](#)

Url for dataset: [CL653Courseworkproject/smoke_detection_iot.csv.zip at main · BSAIMUKESHREDD/CL653Courseworkproject · GitHub](#)

Access point url: