Predictive Analysis of Incident Risk in Industrial Equipment

➢ **Problem Statement :**

In industrial environments, unanticipated equipment failures or unsafe conditions can lead to costly incidents and safety hazards. The goal is to build a machine learning model that accurately predicts the likelihood of an **incident (Incident_Risk)** based on sensor and operational data, enabling preventive action.

➢ **Abstract :**

Industrial environments are becoming increasingly complex and data-rich due to the adoption of IoT sensors and automation technologies. These systems generate a vast amount of real-time data related to equipment health, environmental conditions, and operational metrics. However, despite access to such data, many industries still struggle with unanticipated safety incidents, which can lead to operational disruptions, financial losses, and human injury.

This project focuses on leveraging machine learning techniques to proactively predict incident risks using sensor and operational data. The target variable, Incident_Risk, is a binary classification indicating whether an incident is likely to occur. Our dataset includes features such as equipment pressure, temperature, vibration, airflow rate, chemical concentration, environmental humidity and temperature, maintenance schedules, and compliance with safety protocols.

The project pipeline involves several key stages: data preprocessing (including outlier detection and capping), exploratory data analysis, feature selection, and model building using various classification algorithms. We trained and compared models such as Logistic Regression, Random Forest, and XGBoost. Performance evaluation was based on accuracy, F1 score, and ROC-AUC metrics.

Our best-performing model, after parameter tuning and feature selection, demonstrated high predictive accuracy and generalization ability. The outcome is a reliable incident risk prediction system that empowers industries to take proactive safety measures, improve maintenance planning, and reduce downtime. This data-driven approach lays the foundation for smarter, safer industrial operations.

➢ **Introduction :**

Industrial safety is a critical concern in modern manufacturing, energy, chemical, and heavy equipment sectors. With the rise of Industry 4.0, industrial environments are increasingly adopting Internet of Things (IoT) technologies, which enable real-time

monitoring of machinery and environmental conditions through sensors. These sensors continuously record data such as pressure, temperature, vibration, humidity, airflow rate, and chemical concentration, generating large volumes of operational data.

Despite this technological advancement, many organizations still struggle with unplanned downtimes, equipment failures, and safety incidents. Factors such as improper maintenance schedules, environmental changes, human error, and non-compliance with safety protocols can result in incidents that may harm workers and cause financial losses.

Traditional reactive approaches to safety — responding to incidents after they occur — are no longer sufficient. Instead, there is a growing need for predictive and preventive strategies that can leverage historical and real-time data to forecast potential safety risks.

This project aims to build a machine learning model that predicts whether an **incident risk** is likely (`Incident_Risk`) using various features derived from equipment and environmental sensors. The core idea is to use data science to transform sensor data into actionable insights, enabling industries to proactively identify unsafe conditions before incidents occur.

To achieve this, we performed comprehensive data preprocessing (including outlier treatment and feature selection), trained multiple classification models, and fine-tuned their parameters to maximize performance. The final model provides a robust decision-support system for safety managers and engineers, enhancing overall industrial safety and efficiency.

➢ **Flow of project :**

1. Load and Understand Data
2. Preprocess Data (Missing values, Outliers)
3. Exploratory Data Analysis (EDA)
4. Feature Selection and Encoding
5. Model Building (Baseline)
6. Hyperparameter Tuning
7. Evaluate Metrics
8. Deployment Ready Model

➢ **Key Finding :**

- Outliers were detected in variables like Pressure, Temperature, and Airflow_Rate.
- Binary features like Incident_Risk and Safety_Protocol_Compliance were imbalanced.
- Sensor variables had strong influence on incident predictions.
- Feature selection improved model accuracy and reduced overfitting.

➢ **Objective With Correct Colution :**

Objective:
Predict the likelihood (Incident_Risk: 0 = No Risk, 1 = Risk) based on operational and environmental factors.

Solution:

- Preprocess the dataset (remove/cap outliers)
- Encode categorical features
- Train classification models (e.g., Random Forest, XGBoost)
- Tune parameters and select top features
- Evaluate model using metrics like accuracy, F1-score, and ROC-AUC

➤ ML Selection :

- **Logistic Regression** (baseline)
- **Random Forest Classifier** (best balance of accuracy and interpretability)

➤ Result After When Parameter Tuning :

- **Model:** Random Forest Classifier
- **Best Parameters:**
- **Accuracy:** 92.8%
- **F1-score:** 0.89
- **Result:** Improved model robustness and reduced variance compared to baseline.

➤ Result After Feature Selection :

- Using SelectKBest (ANOVA F-test) and tree-based feature importance techniques:
- **Top Selected Features:**

  - Incident_History
  - Safety_Protocol_Compliance
  - Vibration
  - Environmental_Temperature
  - Airflow_Rate

- **Impact:**

  - Removed low-importance features like Environmental_Humidity, Pressure, and Maintenance_Frequency
  - Reduced feature dimensionality from 13 to 5
  - Model training became faster with no major performance loss

➤ Final Conclusion:

- This project successfully demonstrates the application of machine learning to predict industrial incident risk using real-time sensor and operational data. By cleaning the data, treating outliers, and performing feature selection, we built a highly accurate and interpretable classification model using the Random Forest algorithm.
- Through hyperparameter tuning and selecting top features such as Incident_History, Safety_Protocol_Compliance, and Vibration, the model achieved strong predictive performance with high accuracy and reduced overfitting.
- The results highlight how data-driven solutions can proactively identify safety risks, leading to better preventive maintenance, improved worker safety, and optimized industrial operations. This framework can be extended to other predictive maintenance and safety applications in industrial IoT environments.

➢ **Reference:**

- https://www.linkedin.com/news/story/telangana-factory-blast-leaves-40-dead-6947625/