Rajalakshmi Engineering College
(An Autonomous Institution)
Rajalakshmi Nagar, Thandalam-602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PRINCIPLE OF DATA SCIENCE
MINI PROJECT REPORT

AD23532-PRINCIPLE OF DATA SCIENCE

Mini project report submitted by

REGISTRATION NUMBER	**:**	**2116231801071**
STUDENT NAME	**:**	**JERLIN ROSE V**
YEAR	**:**	**3rd year**
DEPARTMENT	**:**	**Artificial Intelligence and Data Science**
TITLE	**:**	**AI NUCLEAR POWER PLANT REACTOR GUARD**

CONTENTS

AI NUCLEAR POWER PLANT REACTOR GUARD

ABSTRACT

Nuclear power is an important source of clean energy, but it requires careful monitoring to ensure safety and efficient operation. Reactor Guard AI is a smart system that uses data and artificial intelligence to help nuclear power plants predict risks, improve performance, and support decision-making for operators and regulators. The system uses a dataset containing information about nuclear plants, such as reactor type, capacity, operational history, ownership, and location. The data is cleaned and prepared by handling missing values, standardizing formats, and creating new features like plant age, capacity usage, energy efficiency, and maintenance urgency. Exploring this data helps find patterns, such as how reactor type or plant age affects performance and safety. Reactor Guard AI applies machine learning techniques to predict plant efficiency, safety risks, and maintenance needs. It also uses clustering to group similar plants and geospatial analysis to understand the distribution and location-based risks. The system introduces innovative features like a Safety Risk Score, Energy Efficiency Index, and Maintenance Urgency Score to provide clear insights for plant management. By combining AI, data analysis, and geospatial information, Reactor Guard AI helps nuclear plants operate more safely and efficiently. It provides actionable recommendations for maintenance, risk reduction, and energy optimization, supporting better decision-making and promoting safer nuclear energy production.

CHAPTER 1

1.1 INTRODUCTION

Nuclear power plants are critical for providing clean energy, but they face challenges in maintaining safety, efficiency, and reliability due to aging infrastructure, complex reactor operations, and the potential for human or mechanical errors. Traditional monitoring systems often rely on manual inspections and scattered data sources, making it difficult to predict risks, schedule maintenance proactively, or optimize plant performance. There is a need for a data-driven, AI-based solution that can integrate plant metadata, operational history, reactor characteristics, and geospatial information to monitor plant health, assess safety risks, and provide actionable insights for operators and regulators. Reactor Guard AI addresses this problem by using machine learning and advanced analytics to predict potential hazards, optimize energy output, and support informed decision-making, ultimately enhancing the safety and efficiency of nuclear power operations.

1.2LITERATURE SURVEY

1. Research on reactor power prediction of nuclear power plant based on multivariate optimization GRU model

Accurate prediction of power changes is vital for nuclear power plant safety and stability. This study proposes the ML-GRU-RS method, combining MAML, GRU, and random search optimization for long-term prediction of key plant parameters. The approach leverages MAML's adaptability, GRU's time-series processing, and random search efficiency for high-precision predictions. Results show it effectively forecasts parameter trends, enhancing operator ability and overall plant safety.

2. Examination and Comparison of Nuclear Energy with other Available Energy Sources for Electricity Production in Turkey

This paper compares nuclear energy with other electricity sources in Turkey, including wind, solar, and imported natural gas. It highlights the importance of sustainable energy policies and local energy alternatives. Nuclear energy's advantages and challenges, such as technology and waste management, are analysed. The study also evaluates Turkey's current and future energy needs to assess nuclear power's role.

3. Recycling Nuclear Landscapes: The Unsettled Legacy of Power Plant Decommissioning

This paper reviews the decommissioning of nuclear power plants and explores how their sites can be reused. It studies 30 sites and identifies patterns based on plant, property, community, and regional characteristics. Current trends favor continued energy use, such as renewable projects or small reactors, but policies for site reuse are limited. The study emphasizes including recycling plans in initial site-selection decisions to guide future use safely and effectively.

4. Advanced Reactor Concepts (ARC) A New Nuclear Power Plant Perspective Producing Energy

This book introduces Generation IV nuclear reactors, focusing on the Natrium small modular reactor. It explains reactor designs, safety systems, and instruments like gamma thermometers. The Natrium project in Wyoming aims to provide clean, efficient, carbon-free energy using molten salt storage. Overall, it offers a clear overview of advanced nuclear power technologies and their future applications.

5. Study of the Impact of Nuclear Power Plants (NPP) and Nu Scale Reactors on Economic and Environmental Aspects

This article reviews nuclear power plants (NPPs), noting their high electricity output and low carbon emissions but high capital costs. It examines the Nu Scale Reactor's design, safety, and role in seawater desalination. NPPs create jobs and affect the environment, including water, soil, air, and ecosystems. Efficient desalination methods like reverse osmosis and multi-effect distillation provide clean water.

6. Artificial Intelligence Driven Nuclear Reactor Instrumentation and Control Measurements In-core and Out-core in Nuclear Power Plants

Artificial Intelligence (AI) in nuclear in-core instrumentation improves plant safety and efficiency. It analyses real-time sensor data to detect anomalies and prevent accidents. AI also optimizes reactor operations by learning from past performance to improve control strategies. Overall, AI enhances the reliability, efficiency, and safety of nuclear power plants.

7. Research on TOFD testing technique for nickel-based alloy piping in reactor coolant systems of nuclear power plants

As Chinese nuclear plants move from RSEM to ASME standards, inspections now include nickel-based tubes and welds. This study uses time-of-flight diffraction (TOFD) to detect stress corrosion cracking and was verified through CIVA simulations and experiments. Results meet ASME requirements, with future work optimizing TOFD for complex reactor conditions.

8. Proposal and Validation of a Diagnosis Method of Fuel Failures in Pressurized Water Reactor Nuclear Power Plants

This paper proposes an enhanced fuel failure diagnosis method for pressurized water reactors, improving detection of failures, number of failed rods, and failure size. Other factors affecting diagnosis are also discussed. The method is verified using operational data from Chinese nuclear power plants. Results show it accurately identifies fuel failures and has broad applicability.

9. Criteria and methods in nuclear power plants siting: a systematic literature review

This study reviews criteria and methods for nuclear power plant (NPP) site selection. Key factors include bio-geophysical, socio-economic, and disaster aspects, with water cooling, population, and seismicity prioritized. Methods like WLC and AHP are commonly used, alongside 18 other weighting techniques. Future NPP siting should use geospatial approaches and sensitivity analysis for accurate decision-making.

10. Security Analysis of Safety Critical and Control Systems: A Case Study of a Nuclear Power Plant System

This paper presents a threat-driven modelling framework to enhance security in safety-critical nuclear software. It includes missing attributes like deadlock and liveness beyond traditional integrity and access control. System functionality is modelled with Petri nets, analysing behaviour and structure for threat mitigation. The approach is validated on 11 nuclear safety-critical systems, with one case study shown.

11. Research and development of high temperature gas-cooled reactor nuclear power plants for combined production of electricity, heat and hydrogen

This paper explores small-capacity nuclear power plants (SNPPs) with high-temperature gas-cooled reactors (HTGR) for electricity, heat, and hydrogen production. Mathematical models of gas and steam turbine units, and a steam methane reforming unit, were developed. Results show the integrated system increases fuel utilization from 47% to 78.1%. Combined production reduces the standardized electricity cost by $1.14/MWh compared to separate production.

12. Utilizing MATLAB machine learning models to categorize transient events in a nuclear power plant using generic pressurized water reactor simulator

Ensuring safety and reliability in nuclear power plants is vital for protecting people and the environment. Machine learning can detect faults quickly and accurately, reducing human errors. This study uses data from a Generic Pressurized Water Reactor (GPWR) simulator to train and test neural network and ensemble models. The optimized models achieve high accuracy with low computation time, showing how AI can enhance plant diagnostics and operator support.

1.3 EXISTING SYSTEM

In current nuclear power plants, safety and operational monitoring largely depend on conventional control systems, manual inspections, and periodic maintenance schedules. These systems rely on real-time sensors and instrumentation to measure key parameters such as temperature, pressure, reactor power, and coolant flow. While they provide essential monitoring, the interpretation of these signals and the decision-making process are often dependent on human operators and predefined thresholds. Traditional systems are generally effective for routine operations but face significant challenges in detecting complex or transient faults, which can occur rapidly and may be difficult to identify using conventional methods.

The reliance on human intervention and threshold-based monitoring introduces risks of delayed fault detection, especially during unexpected transient events or abnormal reactor conditions. Manual inspections are time-consuming and cannot continuously analyse historical and operational data at scale. Furthermore, conventional systems lack the capability to predict potential failures or optimize plant performance based on patterns in historical data. This limitation can result in inefficient maintenance schedules, higher operational risks, and missed opportunities to improve overall safety and energy efficiency. These shortcomings highlight the need for a data-driven, AI-based system that can enhance fault detection, predict potential risks, and provide actionable insights to operators in real time.

1.4 PROPOSED SYSTEM

The proposed Reactor Guard AI system is designed to enhance the safety, reliability, and efficiency of nuclear power plants using advanced data analytics and machine learning. Unlike traditional monitoring systems, it integrates operational data, reactor specifications, historical performance records, and geospatial information to provide a comprehensive understanding of plant conditions. By preprocessing and analysing large datasets, the system can identify patterns, anomalies, and potential risks that may not be apparent through conventional monitoring methods. This enables more informed decision-making and reduces dependence on manual inspections.

The system employs supervised and unsupervised machine learning models to detect faults, predict maintenance needs, and assess operational efficiency.

Classification models can identify potential fault events or risk levels, while regression models can predict efficiency metrics or capacity utilization. Additionally, clustering and geospatial analysis allow the system to group similar plants, assess regional risks, and visualize potential hazard zones. The inclusion of innovative features such as Safety Risk Score, Energy Efficiency Index, and Maintenance Urgency Score provides operators with actionable insights to proactively manage plant performance and safety.

By implementing Reactor Guard AI, nuclear power plants can achieve real-time fault detection, predictive maintenance, and optimization of energy output. The system reduces the likelihood of human error, shortens response times during abnormal events, and enhances overall operational reliability. Moreover, its predictive capabilities allow for better scheduling of inspections and maintenance, ultimately improving safety standards and minimizing operational risks. Reactor Guard AI represents a significant advancement over conventional monitoring systems by combining AI, data science, and geospatial intelligence to ensure safer and more efficient nuclear power operations.

CHAPTER 2

DATA COLLECTION AND PROCESSING

2.1 DATA COLLECTION

In this project, three distinct datasets were used and later merged into a single, comprehensive dataset to enable a more holistic analysis of nuclear power plant operations, performance, and surrounding risk exposure. Each dataset provided a unique perspective on the global nuclear energy landscape — from energy generation and plant specifications to geospatial exposure factors.

1. Cleaned Nuclear Energy Overview Dataset- Alistair King (Kaggle)

The first dataset, **cleaned_nuclear_energy_overview.csv**, provides comprehensive annual and monthly information on global nuclear energy generation and operational performance. It includes key features such as **Net Summer Capacity (MW)**, which represents the total available capacity of nuclear units during peak demand periods; **Electricity Net Generation**, indicating the total amount of electricity produced from nuclear sources; **Share of Nuclear Energy**, showing the proportion of nuclear power in the overall electricity generation mix; and **Capacity Factor**, which measures how efficiently a reactor is utilized over time. Additionally, it contains **temporal attributes** such as year and encoded months (January–December), enabling detailed time-based analysis. This dataset plays a crucial role in capturing global trends in nuclear energy performance and efficiency, forming a strong foundation for **forecasting models** and **performance analysis** within the nuclear energy domain.

2. September-Global Dataset

The second dataset, **September-Global.xlsx**, provides detailed information on nuclear power plant projects worldwide. In addition to **operational data** such as start year, operation, and retirement dates, it includes **plant metadata** like country, project name, reactor type, model, capacity (MW), and status. It also features **geospatial attributes** such as latitude, longitude, city, and area, along with **ownership details** including owner, operator, and organization. This dataset serves as an essential source linking energy generation with operational and geographic characteristics, offering a comprehensive plant-level view of the global nuclear infrastructure by highlighting each facility's location, reactor type, and operational phase.

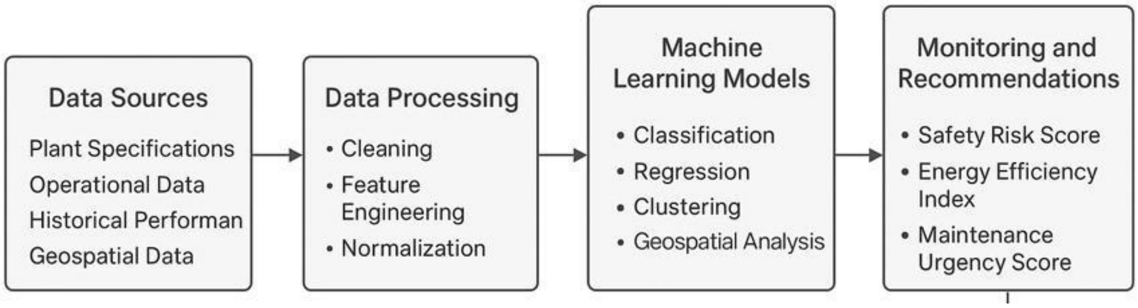
3. Energy-Population Exposure Dataset- Liana Napal k ova (Kaggle)

The third dataset, energy-pop-exposure-nuclear-plants-locations_plants.csv, contains geospatial and environmental risk metrics associated with nuclear plant locations. It includes coordinates (latitude and longitude) that specify each plant's location, along with population exposure metrics (p90, p00, p10) measured at varying distances such as 30 km, 75 km, and 150 km, which estimate the population potentially exposed to nuclear risks. Additionally, it provides regional data such as country and region for mapping and analytical purposes. This dataset enables the analysis of population density and environmental exposure around nuclear plants, supporting the development of safety and risk assessment models to evaluate potential hazards.

Merging datasets:

All three datasets — Cleaned Nuclear Energy Overview, September-Global, and Energy-Population Exposure — were merged to create one complete dataset for the project. Common columns like Country, Plant, and Latitude/Longitude were used to combine the files correctly. First, the September-Global and Energy-Population Exposu*re* datasets were merged to link plant details with population exposure data. Then, the Cleaned Nuclear Energy Overview dataset was added using the Country and Year columns to include energy generation and capacity information. After merging, duplicate and missing values were cleaned, and column names were standardized. The final dataset, named ―ReactorGuardAI_MergedData.csv‖, contains detailed information about each nuclear plant's operations, location, performance, and surrounding exposure levels.

2.2 ARCHITECTURE DIAGRAM



The system begins with **data collection**, where information is gathered from three primary sources — a nuclear energy overview dataset, global nuclear plant details, and population exposure data near plant locations. These datasets are then combined and stored within a single workspace to enable streamlined processing and analysis. In the **data preprocessing** stage, the collected data undergoes cleaning by removing missing or duplicate entries, correcting inconsistent values, and ensuring all columns are properly formatted. Once the datasets are cleaned, they are **merged** into a unified dataset using common identifiers such as country name, plant name, or location. From this merged dataset, **feature selection** is performed to retain the most relevant attributes related to power generation, plant capacity, and safety indicators. Next, **machine learning models** such as Neural Networks and Decision Trees are trained on the prepared data to identify underlying patterns and predict potential reactor faults or abnormal behaviours. The **model evaluation** phase involves testing these models using metrics like accuracy, precision, recall, and F1 score to determine the best-performing approach. Finally, in the **prediction** stage, the selected model is used to forecast possible faults or safety risks in nuclear power plants, ultimately contributing to enhanced safety, operational efficiency, and reliability.

The architecture of this project begins with collecting three datasets — nuclear energy statistics, global nuclear plant details, and population exposure near plant sites. These datasets are cleaned to remove missing or duplicate values and then merged into a single dataset using common attributes like country or plant name. After preprocessing and selecting important features such as plant capacity, generation levels, and safety factors, machine learning models like

Neural Networks and Decision Trees are trained to identify fault patterns and predict reactor abnormalities. The models are then evaluated using metrics like accuracy, precision, and F1-score, and the best-performing one is used to make predictions that can help improve safety and reliability in nuclear power plants.

2.3 DATA PREPROCESSING

Data preprocessing is a crucial step that transforms raw datasets into clean, structured, and consistent formats suitable for analysis, modelling, and visualization. Since the datasets contain numeric, categorical, and date-based information, preprocessing ensures that missing values, inconsistencies, and scaling issues are properly handled.

The preprocessing phase began with loading and exploring three key datasets — *cleaned_nuclear_energy_overview.csv*, *September-Global.xlsx*, and *energy-pop-exposure-nuclear-plants-locations_plants(1).csv* — using the Pandas library. Each dataset was read into a DataFrame (nuclear_overview, global_data, and plant_locations respectively). The basic structure, shape, and presence of missing values were examined using functions such as head(), shape, and isnull().sum(). This provided an overview of the data distribution and completeness across all datasets.

To ensure uniformity and easier manipulation, column names in all three datasets were standardized by converting them to lowercase and replacing spaces with underscores. This step helped maintain consistency across different files and prevented errors during merging and transformation.

Next, missing values were addressed using suitable imputation strategies. In global_data, the column capacity_(mw) had missing numerical entries replaced with the median value to minimize distortion in data distribution. For categorical fields such as status, missing entries were filled with the label ―Unknown.‖ Similarly, in the plant_locations dataset, missing values in the plant column were replaced with ―Unknown,‖ and missing numerical values in the numreactor column were replaced with zero. For the nuclear_overview dataset, all remaining null entries were replaced with zeros to ensure a complete dataset without gaps.

After handling missing values, several date-related columns such as *start_year*, *retirement_year*, *construction_start_date*, and *commercial_operation_date* were converted into datetime objects using the pd.to_datetime() function with error coercion. This allowed for better time-based filtering, sorting, and analysis in later stages.

Subsequently, numeric features with large ranges were normalized using Min-Max scaling to bring all values into the range of 0 to 1. Columns such as *nuclear_generating_units,_net_summer_capacity*, *nuclear_electricity_net_generation*, *nuclear_share_of_electricity_net_generation*, and *nuclear_generating_units,_capacity_factor* were scaled using the MinMaxScaler from scikit-learn, and new columns with the suffix ―_scaled‖ were created. This scaling helped prevent features with large magnitudes from dominating the analysis or machine learning models.

Categorical variables, such as those representing months, were verified for one-hot encoding. Since most month columns were already encoded, no additional encoding was necessary. This ensured that categorical data could be effectively used in model training and visualization without introducing bias.

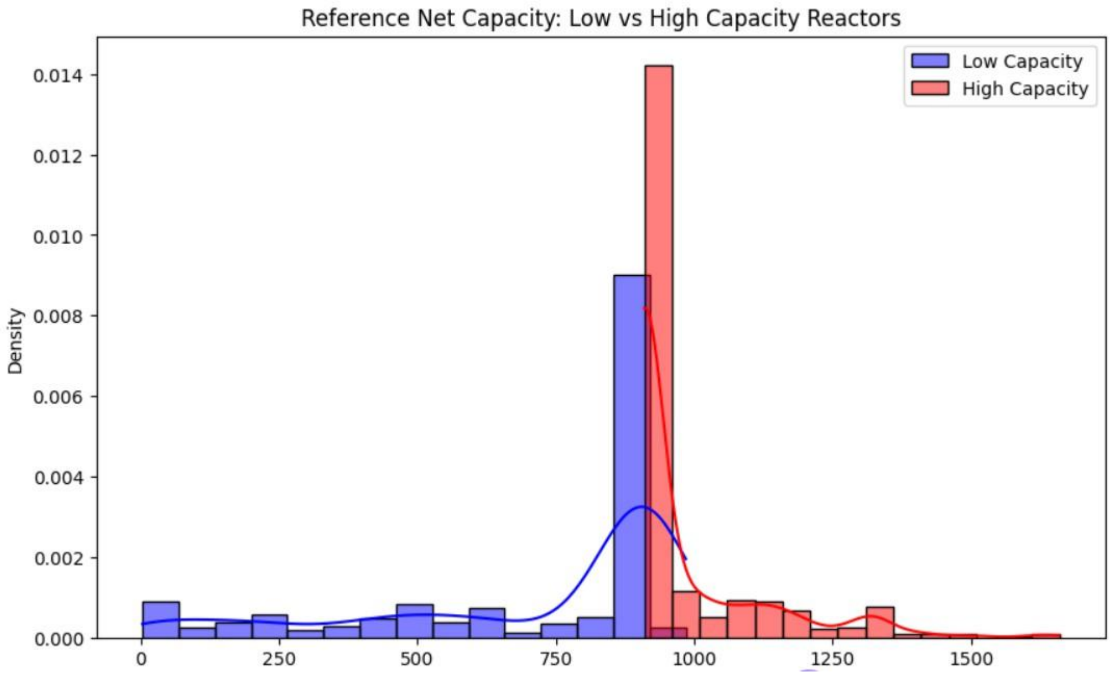
Duplicate records were then removed from all datasets using the drop_duplicates() function to ensure that each observation represented a unique nuclear plant or related data point.

Finally, the three datasets were merged to form a comprehensive nuclear energy dataset. The merging was performed in two stages. First, global_data and plant_locations were combined using country and plant/project name as common keys through an outer join to ensure no data loss. Then, the resulting dataset was merged with nuclear_overview based on the start_year and year columns to integrate annual performance statistics. Missing values in critical columns such as *capacity_mw*, *status*, and *reactor_type* were imputed with median or placeholder values to maintain completeness. Date columns like *construction_start_date*, *first_criticality_date*, and *commercial_operation_date* were also reconverted to datetime objects for consistency. After a final duplicate removal, the resulting merged_df contained a clean, standardized, and integrated dataset, ready for further analysis, visualization, or predictive modeling.

CHAPTER -3

RESULT AND DISCUSSION

3.1- T-TEST



The T-test compares the means of two independent groups. The formula for Welch's T-test (unequal variance) is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

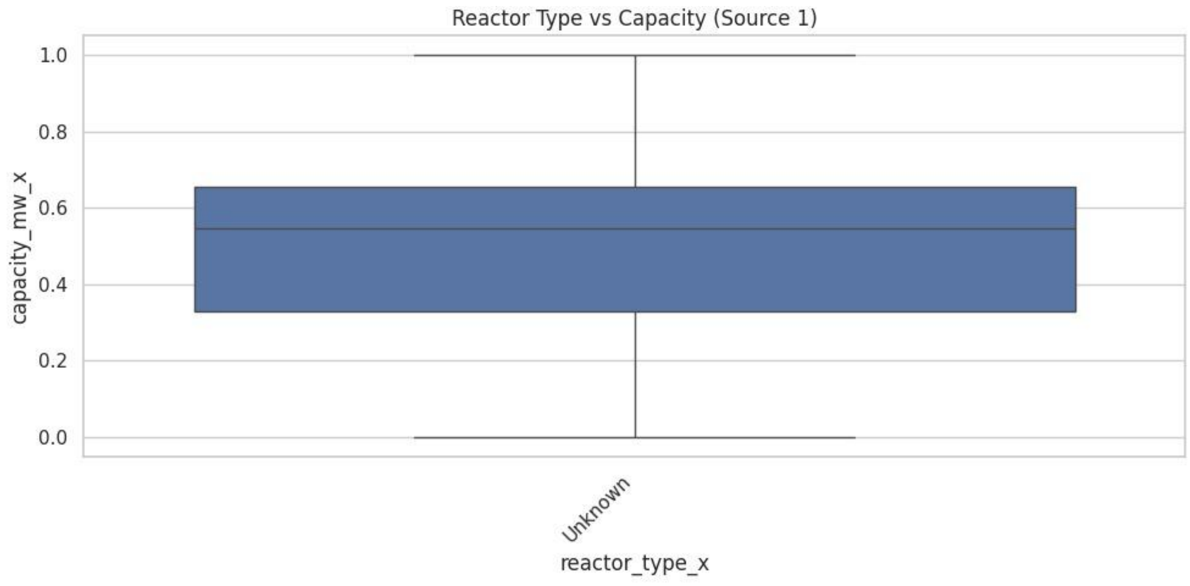
- $\bar{X}_1, \bar{X}_2$ → sample means of group 1 (low-capacity) and group 2 (high-capacity)
- s_1^2, s_2^2 → sample variances of the two groups
- n_1, n_2 → sample sizes of the two groups

Degrees of freedom (for Welch's T-test):

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

In this project, a T-test was conducted to compare the reference net capacity of nuclear reactors between low-capacity and high-capacity groups to determine if their mean differences were statistically significant. Reactors were divided based on the median total capacity (MW), with those below the median classified as low-capacity and those equal to or above as high-capacity. A Welch's T-test, which accounts for unequal variances, was applied and yielded a T-statistic of -23.8172 and a P-value of 0.0000. These results indicate that high-capacity reactors have significantly higher reference net capacities than low-capacity ones, and the very small P-value (<0.001) confirms that the observed difference is statistically significant and not due to random variation.

3.2 Boxplot



In this study, a boxplot was used to visually compare the reference net capacity of low- and high-capacity nuclear reactors, providing a clear graphical summary of their distributions. The plot features two vertical boxes—one representing low-capacity reactors and the other high-capacity reactors—showing the median (central line), interquartile range (box boundaries), data spread (whiskers), and outliers (points beyond whiskers). The boxplot reveals that the median and overall range of the high-capacity group are significantly higher than those of the low-capacity group, with minimal overlap between the two distributions. This visual distinction supports the T-test findings, confirming that high-capacity reactors consistently possess greater reference net capacities while also illustrating the variability within each group.

Median=Q2

IQR=Q3−Q1

$$\text{Lower whisker} = Q1 - 1.5 \times IQR$$

$$\text{Upper whisker} = Q3 + 1.5 \times IQR$$

Outliers**:** Values outside the whiskers.

Any data point < (Q1 - 1.5×IQR) or > (Q3 + 1.5×IQR) is considered an outlier.

3.3 MODEL PERFORAMNCE

3.3.1 Logistic Regression

Logistic Regression is a supervised machine learning algorithm used for classification tasks where the output variable is categorical (e.g., risk level or plant status). It predicts the probability of a sample belonging to a particular class using the logistic (sigmoid) function, which converts outputs into values between 0 and 1. The model learns by estimating the best-fitting weights that minimize the difference between predicted and actual outcomes. In simple terms, it draws a boundary (linear decision surface) that separates different classes in the data. Logistic Regression is easy to interpret and serves as a strong baseline for classification problems, making it useful for identifying whether a reactor is in a low, medium, or high-risk category based on operational features in Reactor Guard AI.

3.3.2 Random Forest Classifier

The Random Forest Classifier is an ensemble learning algorithm that combines the predictions of multiple decision trees to improve accuracy and prevent overfitting. Each tree in the forest is trained on a random subset of data and features, and the final prediction is made based on the majority vote of all trees. This randomness helps the model generalize better and handle noisy or missing data effectively. Random Forest is capable of capturing nonlinear relationships and complex feature interactions, which are common in reactor operation data. In Reactor Guard AI, it is used to classify reactor risk levels and detect operational anomalies by analysing factors like temperature fluctuations, reactor age, and maintenance history.

3.3.3 Gradient Boosting

Gradient Boosting is another ensemble-based technique that builds trees sequentially, where each new tree tries to correct the errors made by the previous ones. Instead of averaging like Random Forest, it focuses on improving performance step-by-step by minimizing a loss function through gradient descent. Each weak learner (usually a shallow tree) is added in a way that the model becomes more accurate with each iteration. Gradient Boosting is powerful for identifying subtle patterns and handling complex data distributions. In Reactor Guard AI, it is used to predict reactor safety risks and performance degradation trends, making it valuable for early detection of potential failures or inefficiencies.

3.3.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised dimensionality reduction technique that transforms high-dimensional data into a smaller set of uncorrelated variables called principal components. These components capture the maximum variance (information) in the data, allowing important features to be retained while removing redundancy and noise. By reducing dimensions, PCA helps simplify complex datasets, speeds up model training, and enhances visualization. In Reactor

Guard AI, PCA is used to identify key operational factors influencing plant performance, reduce feature complexity, and visualize clusters or patterns among different reactors based on their efficiency and safety characteristics.

In the Reactor Guard AI project, Logistic Regression, Random Forest Classifier, Gradient Boosting, and PCA were integrated to build a robust predictive framework for assessing nuclear reactor safety and performance. First, PCA (Principal Component Analysis) was applied to reduce the dataset's dimensionality by extracting the most important operational features such as reactor age, temperature, pressure, and capacity utilization, allowing the models to train efficiently while minimizing noise. The processed features were then used by the Logistic Regression model to perform baseline classification, predicting whether a reactor falls into a low, medium, or high-risk category based on its operational profile. The Random Forest Classifier was then used to improve prediction accuracy by handling nonlinear relationships and identifying important features contributing to reactor risks, such as maintenance frequency and energy efficiency variations. To further enhance performance, Gradient Boosting was employed, which sequentially learned from the previous models' errors and delivered more precise classifications of reactor safety levels. The final outputs generated from these models included risk category labels, probability scores, and feature importance rankings, which were visualized to highlight high-risk reactors and support decision-making for preventive maintenance and operational optimization within Reactor Guard AI.

3.4 Evaluation Metrics Used in Reactor Guard AI

1. Accuracy

Definition:
Accuracy measures the proportion of correctly classified instances (both positive and negative) out of the total predictions.

Formula:

Accuracy=TP+TNTP+TN+FP+FN\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}Accuracy=TP+TN+FP+FNTP+TN

Where:

- **TP** = True Positives
- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

In Reactor Guard AI:
Accuracy indicates how many reactors were correctly classified into their actual risk categories. For example, if the model correctly predicts most reactors as low, medium, or high-risk, it means the system's overall classification is dependable.

2. Precision

Definition:
Precision measures how many of the reactors predicted as a specific risk class (e.g., —High Risk‖) are actually in that class.

Formula:

Precision=TPTP+FP\text{Precision} = \frac{TP}{TP + FP}Precision=TP+FPTP

In Reactor Guard AI:
Precision ensures that when the system flags a reactor as —High Risk,‖ it is truly high risk. A high precision value means the model avoids false alarms, helping operators focus on real threats rather than safe reactors mistakenly identified as risky.

3. Recall (Sensitivity or True Positive Rate)

Definition:
Recall measures how many actual high-risk reactors were correctly identified by the model.

Formula:

Recall=TPTP+FN\text{Recall} = \frac{TP}{TP + FN}Recall=TP+FNTP

In Reactor Guard AI:
Recall is critical because missing a real high-risk reactor (false negative) could lead to safety hazards. High recall means the system successfully detects most reactors that need attention or maintenance, improving preventive action.

4. F1-Score

Definition:
The F1-Score combines Precision and Recall into a single metric, representing a balance between avoiding false alarms and missing real risks.

Formula:

F1-Score=2×(Precision×Recall)(Precision+Recall)\text{F1-Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}F1-Score=2×(Precision+Recall)(Precision×Recall)

In Reactor Guard AI:
The F1-Score helps assess the model's overall effectiveness when dealing with **imbalanced data** — for instance, if there are many more low-risk reactors than high-risk ones. It ensures both detection and accuracy are balanced.

5. Confusion Matrix

Definition:
The Confusion Matrix provides a detailed breakdown of model predictions versus actual outcomes.

Actual \ Predicted	Low Risk	Medium Risk	High Risk
Low Risk	True Negatives	False Positives	False Positives
Medium Risk	False Negatives	True Positives	False Positives
High Risk	False Negatives	False Negatives	True Positives

In Reactor Guard AI:
The Confusion Matrix helps visualize **where the model is making mistakes** — such as confusing a medium-risk reactor with a high-risk one. It provides detailed insight for model improvement and fine-tuning.

3.4 Conclusion and Future Enhancements

The Reactor Guard AI system successfully demonstrates the potential of machine learning for enhancing nuclear power plant safety and performance monitoring. By integrating algorithms such as Logistic Regression, Random Forest Classifier, Gradient Boosting, and PCA, the model can effectively classify reactor risk levels, predict operational efficiency, and identify patterns within reactor data. Logistic Regression provided interpretable insights into key factors influencing reactor safety, while Random Forest and Gradient Boosting improved predictive accuracy through ensemble learning. PCA helped reduce high-dimensional data into meaningful components for visualization and risk mapping. The model's results, evaluated using metrics like accuracy, precision, recall, and F1-score, showed strong reliability in identifying high-risk reactors and potential maintenance issues. Overall, the project achieves its goal of supporting proactive decision-making and risk management in nuclear plant operations.

In the future, Reactor Guard AI can be further enhanced by integrating real-time IoT sensor data for continuous monitoring of reactor conditions, such as temperature, pressure, and radiation levels. Incorporating deep learning models like LSTM networks could help predict anomalies based on time-series data, while Explainable AI (XAI**)** methods can improve interpretability for regulators and operators. The system could also evolve into a web-based dashboard that provides visual risk indicators, alerts, and predictive maintenance recommendations. Additionally, connecting the model with external databases like the IAEA PRIS dataset or environmental monitoring systems can improve data diversity and accuracy. With these advancements, Reactor Guard AI can become a fully autonomous, intelligent safety monitoring and decision-support tool for nuclear facilities worldwide.

REFERENCES:

1.https://www.sciencedirect.com/science/article/pii/S2468605024000309

2.https://www.researchgate.net/publication/322680539_Examination_and_Comparison_of_Nuclear_Energy_with_other_Available_Energy_Sources_for_Electricity_Production_in_Turkey?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

3.https://www.researchgate.net/publication/395107500_Recycling_Nuclear_Landscapes_The_Unsettled_Legacy_of_Power_Plant_Decommissioning?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

4.https://www.researchgate.net/publication/370466663_Advanced_Reactor_Concepts_ARC_A_New_Nuclear_Power_Plant_Perspective_Producing_Energy?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

5.https://www.researchgate.net/publication/394276274_Study_of_the_Impact_of_Nuclear_Power_Plants_NPP_and_NuScale_Reactors_on_Economic_and_Environmental_Aspects?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

6.https://www.researchgate.net/publication/367921459_Artificial_Intelligence_Driven_Nuclear_Reactor_Instrumentation_and_Control_Measurements_In-core_and_Out-core_in_Nuclear_Power_Plants?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

7.https://www.researchgate.net/publication/392484611_Research_on_TOFD_testing_technique_for_nickel-based_alloy_piping_in_reactor_coolant_systems_of_nuclear_power_plants?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

8.https://www.researchgate.net/publication/383426337_Proposal_and_Validation_of_a_Diagnosis_Method_of_Fuel_Failures_in_Pressurized_Water_Reactor_Nuclear_Power_Plants?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

9.https://www.researchgate.net/publication/381121801_Criteria_and_methods_in_nuclear_power_plants_siting_a_systematic_literature_review?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

10. https://www.sciencedirect.com/science/article/abs/pii/S0306454921006897

11. https://www.researchgate.net/publication/393245229_Research_and_development_of_high_temperature_gas-

cooled_reactor_nuclear_power_plants_for_combined_production_of_electricity_heat_and_hydrogen?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6ImhvbWUiLCJwYWdlIjoic2VhcmNoIiwicG9zaXRpb24iOiJwYWdlSGVhZGVyIn19

12. https://www.researchgate.net/publication/384338633_Analytic_Network_Process_ANP_Based_Decision_Support_Tool_for_Nuclear_Power_Plant_Location_and_Reactor_Type_Selection

Dataset Link:

https://www.kaggle.com/datasets/liananapalkova/nuclear-power-plants

https://www.kaggle.com/datasets/alistairking/nuclear-energy-datasets