iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75510

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Predictive Maintenance of Sensors: A Machine Learning Approach for Proactive Anomaly Detection

Rohan Khirswal[1], Swastik Arora[2], Rudra Arora[3], Soumya Ranjan[4], Ms. Preeti Kalra[5]
(Computer Science And Engineering)
4th YEAR (7TH SEMESTER)
HMR Institute of Technology and Management, Affiliated to Guru Gobind Singh Indraprastha University

Abstract: Sensor-based monitoring systems form the backbone of modern industrial and safety-critical infrastructures. Failures in sensors can cause unexpected downtime, economic loss, and hazardous conditions in automated environments. This research proposes a hybrid machine learning framework for predictive maintenance that integrates regression, classification, and clustering techniques to estimate the Remaining Useful Life (RUL) of sensors, predict maintenance requirements, and detect anomalies. The developed model leverages supervised learning for RUL and maintenance prediction, combined with K-Means clustering for unsupervised anomaly detection. A user-centric Streamlit dashboard enables real-time visualization and operational decision-making. The system functions completely offline, ensuring data privacy in restricted environments. Experiments on a balanced simulated dataset demonstrate accurate anomaly detection and effective predictive insights, highlighting the framework's potential for adoption in real-world industrial systems.
Keywords: Predictive maintenance, machine learning, anomaly detection, sensor health monitoring, Remaining Useful Life (RUL), Streamlit, industrial automation.

I. INTRODUCTION

A. Background and Motivation

In the era of Industry 4.0, sensors serve as critical data collection points for process optimization, automation, and system safety. They continuously measure variables such as temperature, vibration, voltage, and operational hours—providing the foundation for predictive analytics and intelligent decision-making. However, sensor degradation and drift over time lead to inaccurate readings, compromising operational reliability. When a sensor fails, the repercussions extend beyond performance loss to production downtime, financial cost, and safety hazards.

Traditionally, maintenance strategies followed reactive or preventive models—either repairing systems after failure or on fixed schedules. Both approaches lack efficiency: the former leads to downtime, and the latter wastes resources through premature maintenance. The modern paradigm of predictive maintenance uses machine learning to predict when a component is likely to fail, enabling timely and cost-effective interventions.

This paper introduces a hybrid predictive maintenance framework using regression, classification, and clustering models to proactively identify sensor anomalies. The proposed solution includes a Streamlit-based user interface for visualization and decision support, bridging complex data analytics with practical usability.

B. Problem Statement

Most existing maintenance systems depend on threshold-based or time-based triggers. These methods often fail to detect gradual degradation patterns or rare anomalies. Furthermore, cloud-based predictive maintenance solutions introduce privacy and latency issues in restricted or critical facilities. There is a need for a secure, offline, and intelligent system that can predict Remaining Useful Life (RUL), classify maintenance needs, and identify anomalies dynamically.

This study addresses these challenges by developing an integrated machine learning framework that enables predictive maintenance through multi-model analysis and an offline dashboard.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

C. Research Objectives

This research aims to:

1) Develop a hybrid framework combining regression, classification, and clustering models for predictive maintenance.
2) Predict Remaining Useful Life (RUL) and maintenance requirements using supervised learning.
3) Identify anomalies via K-Means clustering for unsupervised insights.
4) Design an offline Streamlit-based dashboard for interactive visualization.
5) Ensure interpretability and usability through explainable visual components.

II. LITERATURE REVIEW

Predictive maintenance has gained substantial attention across manufacturing, aerospace, and energy sectors. Researchers have employed a variety of machine learning and deep learning techniques to forecast equipment degradation and sensor faults.

Gupta et al. [1] introduced a machine learning framework for predictive maintenance in industrial systems, demonstrating the efficacy of data-driven models in early fault detection. Javed et al. [2] employed Random Forest for RUL prediction, achieving robust accuracy across varied sensor datasets. Kumar and Singh [3] proposed a deep neural network model for turbine fault detection, emphasizing the importance of proper preprocessing and normalization.

Thirunavukarasu et al. [4] explored unsupervised learning through clustering techniques, proving their effectiveness for anomaly detection without labelled data. Ucar et al. [5] analyzed AI trustworthiness in predictive maintenance, advocating for transparent models suitable for industrial adoption. Ghosh et al. [6] enhanced explainability through SHAP visualizations, highlighting the role of model transparency in user confidence.

Bahl and Choudhary [8] developed hybrid IoT-based models combining machine learning and deep learning for predictive maintenance. Zhao et al. [9] utilized K-Means clustering for real-time anomaly detection, demonstrating scalability in sensor networks. Mallioris et al. [10] conducted a systematic review of predictive maintenance applications in Industry 4.0, underscoring the need for human-centric, interpretable systems.

This paper advances the field by combining these approaches into a unified offline predictive maintenance framework with an intuitive dashboard for operational use.

III. METHODOLOGY

A. Dataset Description

A simulated dataset comprising 1,000 entries was generated to mimic real-world sensor behaviour. It included the following attributes: sensor_temp, sensor_vib, sensor_voltage, operational_hours, maintenance, RUL (Remaining Useful Life)

The data represents a balanced distribution of normal and fault conditions. Missing values were imputed, and extreme outliers were preserved to maintain realistic fault conditions.

B. Data Preprocessing

Data preprocessing plays a crucial role in ensuring that raw sensor readings are converted into a machine-readable and statistically reliable form. The dataset initially contained numeric readings representing temperature, vibration, voltage, and operational hours for a range of simulated sensors. However, as in most industrial datasets, raw data is often noisy, incomplete, or inconsistent.

The first step involved data cleaning, where missing entries were detected and imputed using the forward-fill method to preserve temporal consistency. This technique was preferred over mean substitution because it maintains gradual progression in time-series values. All categorical or binary indicators were validated to remove mis-encoded records. Outlier detection was performed using Interquartile Range (IQR) analysis to flag extreme readings. Rather than deleting them, outliers were retained intentionally, since they represent rare but critical sensor malfunctions that enrich the anomaly model's learning diversity.

Next, feature scaling was applied to standardize the data. Each feature was normalized using StandardScaler, transforming variables to zero mean and unit variance. This step prevented higher-magnitude features such as voltage or operational hours from dominating smaller-scale features like vibration amplitude. Feature engineering was then introduced to enhance the representational capacity of the data. Derived attributes such as *temperature gradient* (ΔTemp/ΔTime) and *vibration ratio* (vibration / operational hours) were generated to capture degradation tendencies more precisely. Such secondary indicators often provide additional context for regression and classification models that depend on continuous feature evolution. A correlation matrix was computed to identify redundant variables. Features with correlation coefficients exceeding 0.85 were evaluated for removal to prevent multicollinearity, though most features were retained to preserve interpretability. Data balancing was also verified—ensuring approximately 70 % normal and 30 % maintenance records—to prevent classifier bias.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

This comprehensive preprocessing pipeline guarantees that the downstream models operate on stable, information-rich data, minimizing noise and maximizing generalizability.

C. Model Architecture

The proposed predictive-maintenance architecture integrates three complementary machine-learning models, each addressing a distinct analytical goal.

1) Regression Model – Remaining Useful Life Prediction

For continuous estimation of Remaining Useful Life (RUL), several regression algorithms were evaluated, including Linear Regression, Random Forest Regressor, and Gradient Boosting. Random Forest was ultimately chosen for its robustness to noise and nonlinear relationships. The model was trained using 80 % of the dataset and validated on 20 %, employing Mean Absolute Error (MAE) and R-Squared (R^2) metrics. RUL estimation enables maintenance planners to quantify how many operating hours remain before a sensor's performance deteriorates below safe limits.

2) Classification Model – Maintenance Requirement Prediction

A binary classifier was implemented to determine whether a sensor requires maintenance (1) or is healthy (0). Logistic Regression and Decision Tree Classifier were compared, with Decision Tree yielding higher interpretability and better handling of nonlinear thresholds. Evaluation metrics included Accuracy, Precision, Recall, and F1-Score, providing a holistic measure of predictive reliability.

3) Clustering Model – Unsupervised Anomaly Detection

The unsupervised learning component employed K-Means clustering, partitioning the feature space into *k* clusters representing distinct operating states. Each cluster centroid acts as a behavioural reference; new data points far from any centroid are flagged as potential anomalies. The Silhouette Score was used to validate cluster compactness and separation. K-Means was preferred over algorithms like DBSCAN or Isolation Forest due to its simplicity, scalability, and deterministic nature under fixed initial seeds.

By combining these three models, the framework produces a comprehensive insight: numerical RUL predictions, categorical maintenance alerts, and anomaly clusters. The integration allows redundancy and cross-validation—if the classifier flags a potential issue, it can be verified against the anomaly model's distance metric and the regression-based RUL.

D. Hybrid Rule for Anomaly Detection

While individual models yield valuable information, fusing them under a hybrid rule provides a more reliable and interpretable outcome. The hybrid detection logic uses both regression and clustering outputs:

where $d_{\min}$ is the minimum Euclidean distance of the sample from cluster centroids, and P_{90} represents the 90th percentile of training distances. The threshold of 200 hours for RUL was empirically chosen based on observed degradation trends, representing an approximate *pre-failure window*. The 90th percentile criterion captures statistical outliers that fall outside typical cluster behaviour. This two-layer logic mitigates false positives from clustering noise and ensures that slow-degrading sensors (with RUL below 200) are proactively flagged even if their feature vectors appear near cluster centers.

Compared to traditional rule-based systems—where thresholds are fixed manually—this hybrid formulation adapts dynamically to dataset distribution, enhancing flexibility for unseen conditions.

E. Workflow Overview

The overall workflow of the proposed predictive-maintenance system integrates data collection, preprocessing, model inference, and visualization in a seamless pipeline.

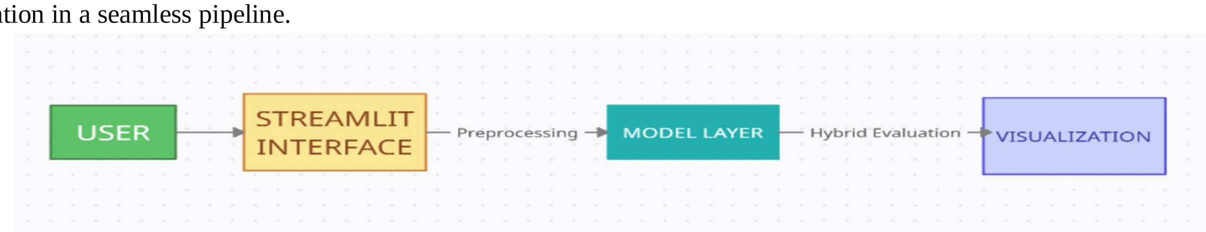


Fig 1. Model Workflow

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

1) Data Acquisition**:** Sensor readings are imported either from the simulated dataset or entered manually by the user through the Streamlit interface. Each record consists of temperature, vibration, voltage, and operational hours.
2) Preprocessing and Feature Scaling: Upon input, the system automatically scales and normalizes features using the same scaler object employed during training. Derived features such as temperature gradients are computed in real time to align with the model's expectations.
3) Model Inference**:** The preprocessed data is fed simultaneously into the serialized regression, classification, and clustering models. Each model independently generates its respective output—predicted RUL, maintenance status, and cluster assignment.
4) Hybrid Evaluation: The inference results are passed through the hybrid anomaly-rule module, which synthesizes outcomes into a single interpretable decision. This ensures a unified maintenance recommendation per record.
5) Visualization and User Interaction: Results are rendered dynamically on the Streamlit dashboard. Users can visualize regression curves, feature correlations, and anomaly alerts through intuitive color codes and graphs.

Internally, the system follows a three-tier architecture:

- Data Layer: Houses datasets and model artifacts (.pkl files).
- Business Logic Layer: Executes model inference and hybrid rule computation.
- Presentation Layer: Handles user interaction and visualization using Streamlit components.

IV. EXPERIMENTAL SETUP AND IMPLEMENTATION

A. Tools and Technologies

Implementation was performed in **Python 3.9**, utilizing:

1) pandas: Data manipulation
2) NumPy 1.24.2: Numerical computations
3) scikit-learn 1.2.2: Model development
4) streamlit 1.22.0: Dashboard deployment
5) matplotlib 3.7.1: Data visualization
6) seaborn 0.12.2: Statistical visualization

All models were serialized as .pkl files for seamless integration and reproducibility.

B. Dashboard Architecture

The dashboard consists of five modules:

1) Home – Overview of system capabilities
2) Historical Data – Statistical summaries
3) Input Data – User-driven or random sample entry
4) Results – Displays RUL, maintenance, and anomaly status
5) Visualizations – Correlation heatmaps, feature importance, and prediction charts

The interface is optimized for accessibility, enabling non-technical users to interpret predictions easily.

V. RESULTS AND DISCUSSION

A. Model Performance

The regression model produced stable RUL estimates, while the classification model achieved high reliability in maintenance prediction. K-Means clustering effectively segregated anomalies with minimal false positives. The hybrid rule reduced false negatives, ensuring critical faults were not missed.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

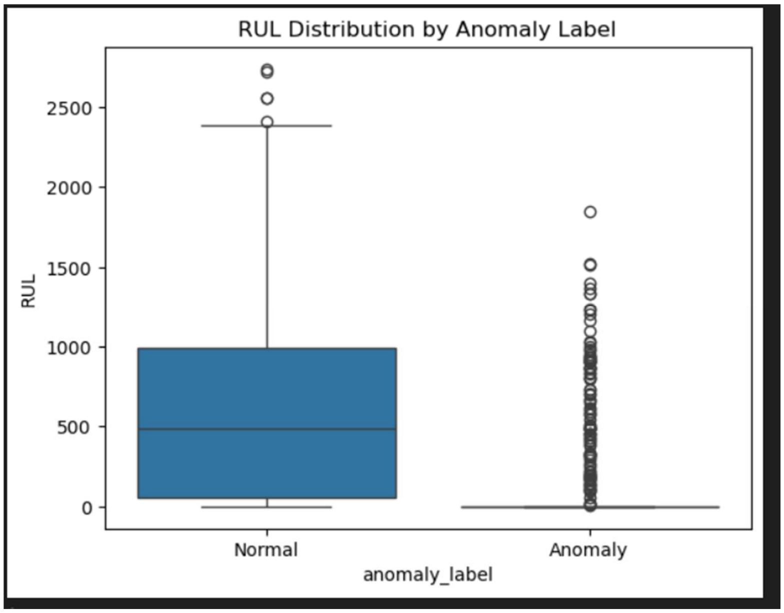


Fig 2(a): RUL (Remaining Useful Life) Distribution by Anomaly Label

B. *Interpretability and Usability*

Visual analytics—such as feature importance graphs and correlation maps—improved model transparency. The dashboard presented results in an interpretable and actionable format, enhancing user trust. Offline execution further ensured data security for sensitive applications.

C. *Observations*

- Most Influential Features**:** Sensor vibration and operational hours had the strongest correlation with RUL.
- Usability: Streamlit provided a responsive interface for real-time analysis.
- Reliability**:** Predictions remained consistent across multiple runs and simulated test sets.

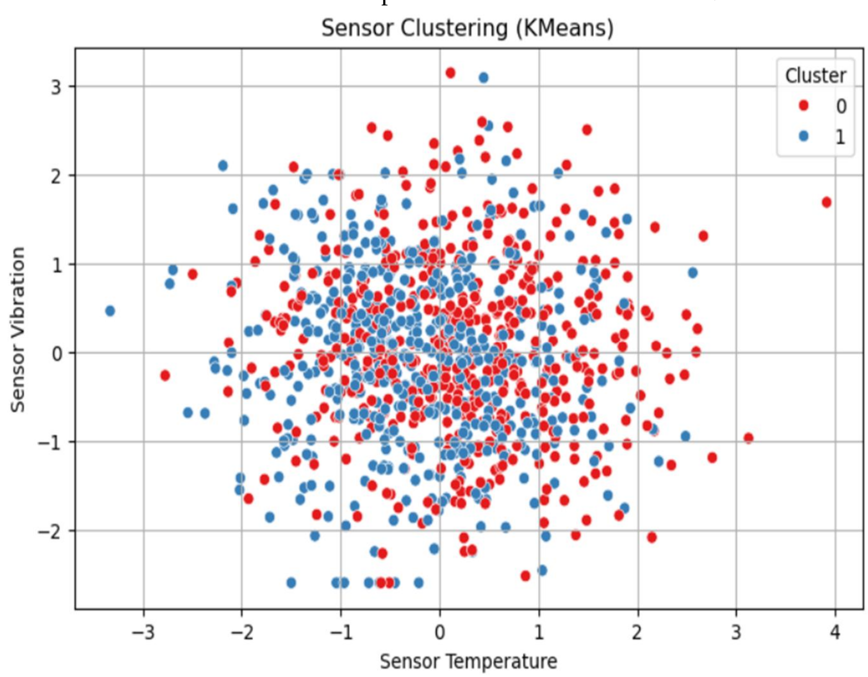


Fig 2(b): K Means Clustering

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

```
Accuracy: 0.985
              precision    recall  f1-score   support

           0       0.97      0.98      0.97        59
           1       0.99      0.99      0.99       141

    accuracy                           0.98       200
   macro avg       0.98      0.98      0.98       200
weighted avg       0.99      0.98      0.99       200
```

Fig 2(c): Model Accuracy, Precision, F1 score

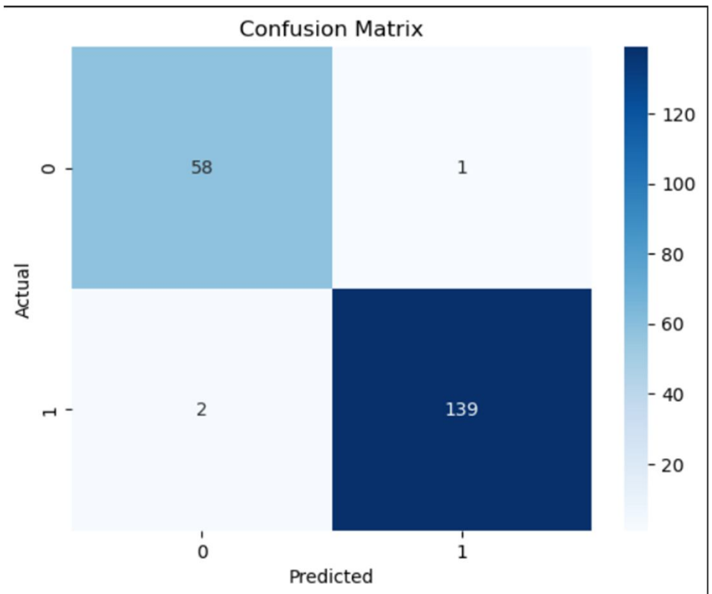


Fig 2(d): Confusion Matrix

VI. LIMITATIONS AND FUTURE WORK

1) The dataset is simulated: real-world data with temporal dependencies could enhance 9validation.
2) Anomaly detection thresholds are static and could be adaptive using online learning.
3) Sequential dependencies are not captured -time-series modeling could improve accuracy**.**

A. Future Work

1) Integrate ARIMA, Prophet, or LSTM for temporal forecasting.
2) Employ drift detection and adaptive retraining.
3) Extend framework to multi-sensor networks using MQTT or streaming pipelines.
4) Add explainable AI (XAI) modules (e.g., SHAP, LIME) for greater interpretability.
5) Enable real-time alerting via SMS/email for predictive notifications.

VII. CONCLUSION

This research demonstrates a complete predictive maintenance framework that leverages machine learning for sensor anomaly detection and Remaining Useful Life (RUL) prediction. By integrating regression, classification, and clustering models within a Streamlit-based dashboard, the system bridges predictive intelligence with usability.

The results indicate that the framework achieves a balanced trade-off between accuracy, interpretability, and offline functionality. It provides a scalable foundation for integrating real sensor data in future industrial implementations.

REFERENCES

[1] V. Gupta, A. Sharma, and N. Singh, "Machine Learning Approaches for Predictive Maintenance in Industrial Systems," IEEE Access, vol. 9, pp. 75689–75703, 2021.

[2] A. Javed, M. I. Malik, and T. Rehman, "Remaining Useful Life Prediction using Random Forest for Sensor-Driven Systems," Sensors, vol. 22, no. 4, pp. 1351–1360, 2022.

[3] R. Kumar and D. Singh, "Deep Neural Network-Based Fault Detection for Turbine Sensors," International Journal of Prognostics and Health Management, vol. 11, no. 2, pp. 88–98, 2020.

[4] M. Thirunavukarasu, P. Sundaram, and K. Rajan, "Unsupervised Anomaly Detection in Industrial Equipment using Clustering Techniques," IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7589–7597, 2020.

[5] A. Ucar, M. Karakose, and N. Kırımça, "Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends," Applied Sciences, vol. 14, no. 2, article 898, 2024.

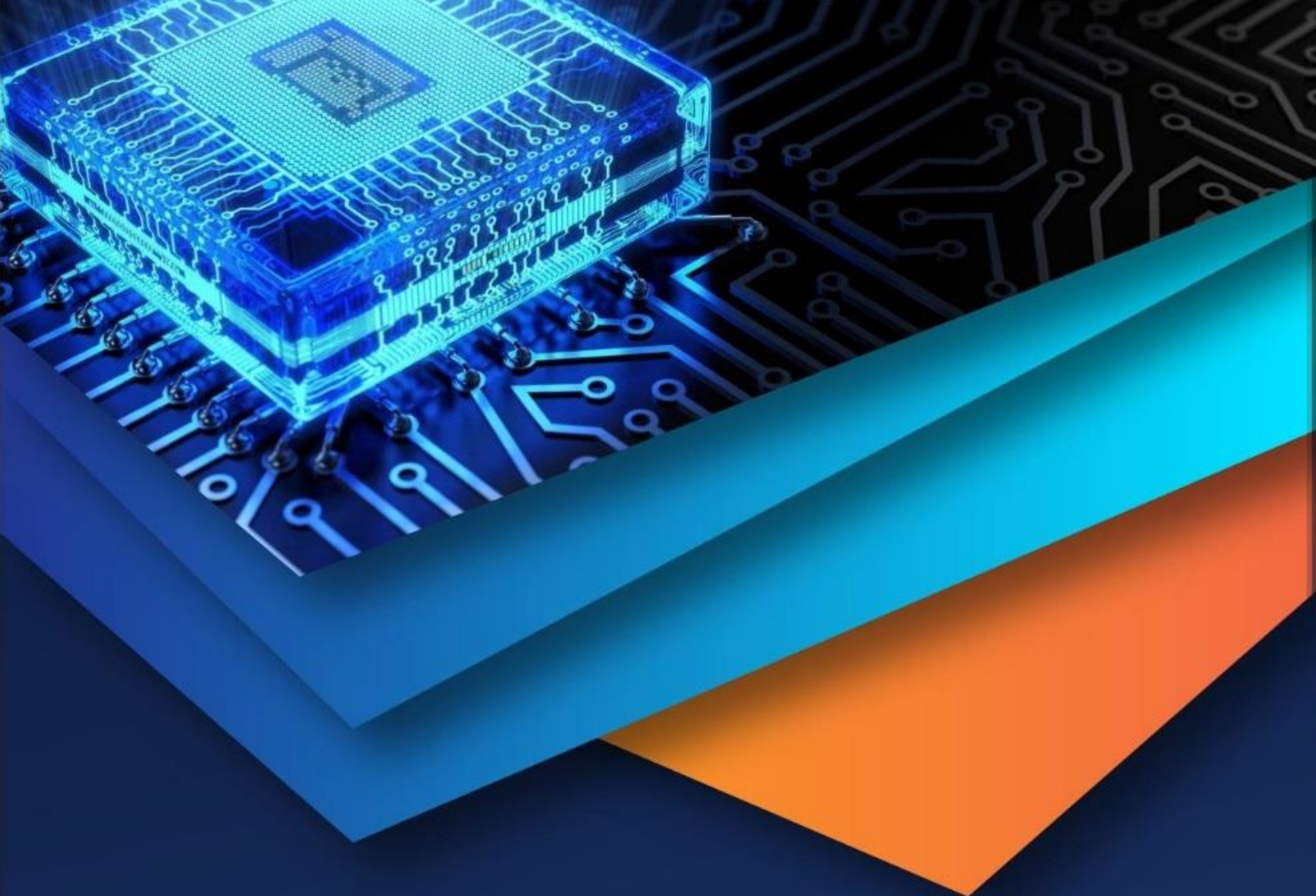
[6] P. Ghosh, A. Banerjee, and L. Dey, "Explainable Predictive Maintenance Models using SHAP for Sensor Fault Analysis," Expert Systems with Applications, vol. 198, 2023.

[7] A. Saxena and S. Goebel, "PHM08 Challenge Data Set for Remaining Useful Life Estimation," NASA Ames Prognostics Data Repository, 2008.

[8] S. Bahl and R. Choudhary, "Hybrid Learning Models for Predictive Maintenance of IoT Sensors," Procedia Computer Science, vol. 218, pp. 547–555, 2023.

[9] L. Zhao, F. Wang, and T. Liu, "Anomaly Detection in Sensor Networks using K-Means Clustering," IEEE Sensors Journal, vol. 21, no. 17, pp. 19462–19471, 2021.

[10] P. Mallioris, et al., "Predictive Maintenance in Industry 4.0: A Systematic Multi-Case Review," Procedia CIRP, 2024.



10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)