Comprehensive Journal of Science
Volume (10), Issue (37), (NOV. 2025)
SICST2025, www.sicst.ly
ISSN: 3014-6266

مجلة العلوم الشاملة
عدد خاص بالمؤتمر الدولي الثالث للعلوم والتقنية
المجلد (10)، العدد (37)، (نوفمبر2025)
ردمد: 3014-6266

Predictive Maintenance of Power Plants in Libya Using Machine Learning Algorithms

Basma Mohammed Albia[1] *, Dr. Juma Ibrahim[2]
Postgraduate Office, Software Development Technology, College of Computer Technology Tripoli (CCTT), Libya

الصيانة التنبؤية لمحطات توليد الطاقة في ليبيا باستخدام خوارزميات تعلم الآلة

بسمة محمد البى[1] *، د. جمعه إبراهيم[2]
مكتب الدراسات العليا، قسم هندسه تطوير البرمجيات، كليه تقنيه الحاسوب طرابلس، ليبيا
bb03002@cctt.edu.ly[1], juma.ibrahim@cctt.edu.ly[2]

Received: 30-09-2025; Revised: 10-10-2025; Accepted: 31-10-2025; Published: 25-11-2025

Abstract

Ensuring the reliability and efficiency of power plants in Libya requires advanced predictive maintenance strategies. Traditional reactive and preventive approaches often fail to prevent unplanned downtime and inefficiencies in modern power systems. This study applies machine learning (ML) techniques, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest, to sensor and historical operational data for predicting maintenance requirements. Comparative evaluation demonstrates that Random Forest achieves superior accuracy (100%) and robustness, while SVM attains high accuracy (93.7%) but shows minor limitations in distinguishing normal and abnormal operational states. The results indicate that ML-based predictive maintenance can significantly reduce unexpected downtime, enhance operational efficiency, and optimize maintenance planning, offering a data-driven alternative to conventional approaches.

Keywords: Predictive Maintenance, Machine Learning, Random Forest.

الملخص

ضمان موثوقية وكفاءة محطات الطاقة في ليبيا يتطلب استراتيجيات متقدمة للصيانة التنبؤية. غالباً ما تفشل الأساليب التقليدية (التفاعلية والوقائية) في منع التوقف غير المخطط له وعدم الكفاءة في أنظمة الطاقة الحديثة. تطبق هذه الدراسة تقنيات تعلم الآلة (ML)، بما في ذلك خوارزمية أقرب الجيران (K-Nearest Neighbors - KNN)، وآلات الدعم الناقل (Support Vector Machines - SVM)، والغابة العشوائية (Random Forest)، على بيانات المستشعرات والبيانات التشغيلية التاريخية للتنبؤ بمتطلبات الصيانة. أظهرت التقييمات المقارنة أن خوارزمية الغابة العشوائية (Random Forest) تحقق دقة متفوقة بلغت 100% وقوة عالية، بينما حققت خوارزمية SVM دقة مرتفعة

بلغت 93.7% ولكنها أظهرت بعض القصور الطفيف في التمييز بين الحالات التشغيلية الطبيعية وغير الطبيعية. تشير النتائج إلى أن الصيانة التنبؤية المعتمدة على تقنيات تعلم الآلة يمكن أن تقلل بشكل كبير من أوقات التوقف غير المتوقعة، وتحسن الكفاءة التشغيلية، وتساعد في تحسين تخطيط الصيانة، مما يوفر بديلاً قائمًا على البيانات مقارنة بالأساليب التقليدية.

الكلمات المفتاحية: الصيانة التنبؤية، تعلم الآلة، الغابة العشوائية.

1. Introduction

The relentless pursuit of operational efficiency and cost reduction has long been a challenge facing industrial and manufacturing strategy (Grace, 2025). Traditional maintenance philosophies, particularly reactive (operation-to-failure) and preventive (preventive) maintenance, have proven increasingly inappropriate in the modern industrial landscape, often leading to unforeseen downtime, over utilization of spare parts, and unnecessary maintenance interventions (Mitchell, et al., 2025). Predictive maintenance represents a transformational maintenance strategy that leverages real-time data from sensors and historical records to estimate equipment condition and accurately predict when maintenance will be performed (Ullah, Hayat, Jhatial, & Yousuf, 2025). Unlike preventive maintenance, which operates on fixed schedule. Traditional maintenance faces many significant challenges, primarily due to the complexity and scope of the systems involved.

Traditional methods often rely on manual inspections based on personnel expertise and scheduled maintenance based on the lifespan of systems and equipment, which can be inefficient and error prone. One of the major difficulties is the inability to accurately predict equipment failures in advance, relying on historical data and predetermined time periods that may not consider real-time operating conditions or subtle changes in equipment behavior. This can lead to unnecessary maintenance or unexpected failures, both of which are financially costly and disruptive (Hu, 2025), (Kusumaningrum, Kurniati, & Santosa, 2021).

Another challenge facing power plant systems is the massive volume of data they generate, which exceeds the capabilities of traditional technologies for effective processing and analysis. This makes it difficult to detect patterns or anomalies that may indicate potential failures. Traditional methods also lack the flexibility to adapt to changing data, limiting the improvement of forecast accuracy over time (Marangis, et al., 2024). Furthermore, traditional predictive maintenance techniques do not integrate well with modern IoT and sensor technologies, which provide real-time monitoring and data collection. This limits the ability to fully leverage the full potential of these technologies to make more accurate and timely maintenance decisions (Khamfoy, Klomwises, & Srianomai, 2025), (Meribout, Saied, & Al Hosani, 2018).

This study presents a novel predictive maintenance framework for power generation systems, overcoming critical limitations of conventional preventive and reactive strategies. These traditional approaches impose significant economic and operational burdens through unplanned downtime and productivity losses. We develop a machine learning model, trained on operational plant data, for strategic fault detection. This model forms the core of a practical predictive maintenance system designed to enhance forecasting accuracy, mitigate unexpected equipment failures, and optimize maintenance scheduling, thereby substantially improving plant reliability and productivity.

2. Related Works

Ahmed Adel Sayed, Taqian Mohammed and Yasmine Ali Shalabidra (2025) studied the Sabiya Steam Power Plant in Kuwait and developed a predictive maintenance framework utilizing machine learning techniques. The study focuses on anticipating maintenance needs by examining critical operating characteristics such as temperature and pressure. operating hours, flow rate, and alarm signals. The methodology relied on preprocessing the data and dividing it into training and test data. Predictive models were then trained using SVM and KNN algorithms, in addition to building a neural network using TensorFlow and comparing them with the improved version, TensorFlow Lite. The results showed that the SVM algorithm achieved the best performance with an accuracy of 0.95, while KNN achieved an accuracy of 0.93. However, the ensemble model (SVM+KNN) did not outperform SVM alone. A comparison between TensorFlow and TensorFlow Lite demonstrated that both achieve the same level of accuracy (0.95), but TensorFlow Lite is distinguished by its small size (25KB) and high speed (0.13ms), making it more suitable for applications on resource-constrained devices. The importance of this study lies in its provision of a practical framework applicable in a real-world operational environment. However, it was limited to a limited set of operational variables and was tested only on a single station, which may limit the possibility of generalizing its results to other stations (Grace, 2025).

Amir Hossein Baradaran (2025) did research to create a model for predicting diagnosis of electric motor health using supervised learning algorithms. The researcher employed a quantitative methodology based on genuine data from 1,050 industrial maintenance reports, which included characteristics including temperature, current intensity, coil resistance, and acoustic condition. Various statistical and intelligent models were used, including Naïve Bayes, logistic regression, k-NN, Random Forest, and SVM (with different kernels). XGBoost, LightGBM, and CatBoost, with performance evaluated using precision, recall, specificity, and F1 coefficient metrics. The results showed that the CatBoost model outperformed other models with an accuracy of 92.86%, outperforming other models due to its ability to handle categorical data efficiently and reduce overgeneralization. The study's advantages include its reliance on actual industrial data, the diversity of the compared models, and the use of comprehensive evaluation criteria, which enhances its reliability and industrial applicability. However, its limitation to one type of engine (160kW capacity) and the sample size limit the generalization of the results to other types of engines or different operating environments (Mitchell, et al., 2025).

B. Dhana Laxmi and G. Sai Chaitanya Kumar (2024) did a study to improve the efficiency of solar power plants through predictive maintenance and defect detection using machine learning algorithms. The researchers used a quantitative methodology employing public data from the Kaggle platform, which contained more than 3,000 samples of sensor data (ambient temperature and module temperature), as well as power. generation data. The data was analyzed using Weka 3.8.6, employing Gaussian Processes, Linear Regression, and SMOreg algorithms, in addition to meta-classifiers such as Bagging, Random Committee, and Random Subspace, and evaluated using correlation coefficient, mean square error (MSE), and root mean square error (RMSE). The results showed that the SMOreg algorithm achieved the highest correlation coefficient (0.9796) and the lowest RMSE (0.0605), while the metaclassifiers were characterized by faster model building (less than one second) compared to other algorithms. The study's advantages include its use of realistic data and its evaluation of a variety of algorithms, including meta-classifiers that strike a balance between accuracy and speed, enhancing its practical applicability in solar power plant management. However, limiting the analysis to two main variables (ambient temperature and module temperature) and not incorporating other environmental variables such as solar radiation or wind speed is a major limitation of the model's comprehensiveness (Ullah, Hayat, Jhatial, & Yousuf, 2025).

Pacheco et al. (2024) presented a study aiming at establishing predictive indicators for power transformers using machine learning algorithms to reduce breakdowns and downtime expenses. The researchers used a quantitative approach with two data sets: chromatographic data from public databases to develop the Chromatographic Assay Indicator (CAI), and operational sensor data from a major electricity company’s SCADA/EMS system to develop the Electrical Failure Risk Indicator (EFRI). The Random Forest approach was chosen after testing it against various algorithms such as SVM and FURIA, as well as data processing techniques (cleaning, SMOTE balance, and feature calibration), and model evaluation using precision, AUC, and recall measures. The results showed that CAI surpassed traditional dissolved gas analysis (DGA) methods, with an accuracy of 92.2% and a 19-percentage point difference from the best conventional approach. EFRI also achieved an accuracy of 95.4% and an AUC of around 89% on test data, indicating a strong capacity to anticipate failures before they occur. The research has several advantages, including the integration of genuine industry data with public databases and the production of two integrated indicators to assist maintenance decision-makers. CAI's limitations include its reliance on a small number of gases available in public databases, as well as the complexity of the data balancing process due to the rarity of failure instances, which may alter the results' applicability to varied operating conditions (Xu, Zhao, & Yu, 2025).

Deshmukh et al. (2024) presented a study aimed at developing an integrated predictive maintenance framework that combines machine learning, Internet of Things (IoT), and self-attention techniques to improve the reliability and efficiency of industrial motors in power plants. The researchers adopted a quantitative methodology based on collecting operational data from three industrial motors across 34 datasets containing more than 10,000 sensory data points (temperature, vibration, current), 500 historical maintenance records, and 100 actual failure cases. Advanced machine learning algorithms were applied, most notably Long Short-Term Memory (LSTM) with self-attention mechanisms and optimization techniques such as Adam Optimization, with performance evaluated using precision, recall, F1, and Mean Squared Logarithmic Error (MSLE) criterion. Results across ten replicated experiments showed an average accuracy of 86.3%, recall of 84.5%, and an F1 coefficient of 86.2%, demonstrating the reliability of the proposed framework in predicting failures early and guiding maintenance strategies. The advantages of the study include the integration of multiple data sources (sensory, historical, and actual faults), the adoption of advanced techniques such as self-attention, which improves prediction accuracy, and the real-time applicability of IoT devices. Limitations include the limited sample size (only three engines) and the high complexity of the models, which may make their practical application on a large industrial scale difficult (Ade & Sheriffdeen, 2024).

Kundacina et al. (2024) presented a study aimed at developing a new framework for detecting operational anomalies in thermal power plants with the aim of enhancing predictive maintenance and reducing unplanned breakdowns. The researchers adopted a quantitative approach using simulation data for a 20 MW Rankine cycle power plant model. More than 600,000 samples were collected, distributed among training, calibration, and test data, which included normal operating conditions and four simulated failure conditions (low pump efficiency, condenser fouling, low cooling water flow rate, and high steam temperature at the boiler outlet). Anomaly detection algorithms (Isolation Forest, One-Class SVM, Local Outlier Factor) were used as a basis, and the Conformal Anomaly Detection (CAD) technique was applied with statistical modifications to adjust the false positive rate. The results showed that the proposed framework reduced the false positive rate to 0.0995 using the Simes modification, while maintaining high sensitivity (recall 0.8913) and an overall accuracy of 89.9%, outperforming traditional methods that suffered from a high false alarm rate (exceeding 65% in the case of Isolation Forest without tuning). A notable advantage of the research is that it presents the first application of CAD in the context of predictive maintenance of thermal power plants with formal statistical guarantees, in addition to its real-time applicability without the need for manual threshold calibration. Limitations include its reliance on simulated data rather than actual industrial data, which may limit the generalizability of the results. Furthermore, the complexity of the statistical framework used may pose challenges for large-scale industrial application (Alburshaid & Al-Alawi, 2024).

Bouaziz, Cherif, and Bacha (2025) conducted a study to improve predictive maintenance scheduling at a 10 MW solar power facility in Tozeur, Tunisia, by merging weather data with hybrid machine learning approaches. The researchers took a quantitative approach, employing historical climate data (2019-2024) acquired by the Solcast API platform, which included variables such as solar radiation, temperature, wind speed, humidity, air pressure, and cloudiness, along with PV system performance data. The study used multiple models: SARIMA to analyze seasonal variations in energy production, Autoencoder to detect environmental anomalies, a hybrid framework combining ARIMA with Isolation Forest to predict anomalies, and finally Random Forest Classifier to generate optimal maintenance schedules based on climate predictions and system status. The results showed that the integrated approach improved maintenance reliability and operational efficiency, accurately predicting seasonal declines and climatic-related failures, and suggested maintenance schedules that reduced system downtime and extended system life. The research's advantages include its comprehensiveness, as it integrates realtime climate analysis with performance data and its application to a real-world case study of a major power plant, enhancing its practical application. However, its heavy reliance on limited historical data from a single source (Solcast) and the lack of performance verification via real-time field operating data remain major limitations that may limit the generalizability of its results (Hu, 2025).

Marangis and colleagues (2024) presented a study aimed at developing an integrated, data-driven routine for fault detection and maintenance prediction in utility-scale photovoltaic (PV) power plants. The researchers adopted a quantitative approach using three years of historical data from a 1.8

MW solar plant in Larissa, Greece, which included electrical (current, voltage, DC/AC power) and climatological (irradiance, temperature) measurements. The framework included three main tools: XGBoost to simulate electrical system performance, One-Class SVM to detect anomalies, and Facebook Prophet to predict performance trends and generate maintenance alerts up to 7 days in advance. The results demonstrated high performance simulation accuracy (nRMSE < 5.4%), a high fault detection efficiency of 96.9%, a predictive sensitivity of 92.9%, and an overall accuracy of 99.4% with limited false alarms. The study's advantages include its integration of statistical modeling and machine learning algorithms, and its reliance on real, large-scale operational data, which enhances its reliability and industrial applicability. However, limitations include its reliance on data from a single site within a specific climate, in addition to the challenges of real-time application, which requires the development of a continuously updated digital twin model (Marangis, et al., 2024).

Remaining Global Research Challenges in Predictive Maintenance

Although significant advances were noticed in terms of predictive maintenance using machine learning through multiple generation systems, research challenges remain globally. First, many studies highly depend on limited datasets from single sites or specific equipment, therefore restricting the general aspect of predictive models to other environments or plants (Marangis, et al., 2024), (Bouaziz, Cherif, & Bacha, 2025). Second, integrating different data sources, including historical maintenance records, environmental variables and IoT sensors, remain an overlooked and complicated process, and this limits the ability to develop entirely adaptive tools. Third, the imbalance in handling class and rare failure events continue to be a significant challenge. It is known that most of the ML models tend to underperform on infrequent but significant operational anomalies (Kusumaningrum, Kurniati, & Santosa, 2021).

Finally, scalability and real-time implementation of predictive maintenance models in industrial environments reflect practical difficulties, and this includes data streaming, model restraining requirements and computational constraints. Identifying those challenges reflect a key direction for future research, to ensure that predictive maintenance systems can achieve accurate, strong and scalable performance on a global scale.

3. Experimental Setup

In the experimental setup, five key operational parameters were considered: temperature, pressure, operating hours, flow rate, and alert signals from the boiler feed pump. The boiler feed pump plays a vital role in power plants by transferring water from the feed tank to the boiler, thereby maintaining appropriate water levels and pressure against internal steam pressure, which is essential for steam generation and power production. These parameters were selected due to their critical impact on system reliability and operational efficiency, as monitoring them allows early detection of deviations from normal conditions that may indicate potential faults or maintenance needs (Marangis, et al., 2024).

For practical implementation, maintenance requirements were categorized into four states: normal, abnormal, early maintenance, and annual maintenance, enabling prioritization of corrective actions based on severity and urgency. The categorical outputs were further encoded using label encoding to facilitate processing by machine learning algorithms (Khamfoy, Klomwises, & Srianomai, 2025).

3.1. Operational Parameters & Data Collection

This research identified five key operational parameters – temperature, pressure, operating hours, flow rate, and boiler feed pump alert signals – as essential for analysis. These variables were continuously tracked through sensor-based monitoring. The dataset contains historical records categorized into four operational states: normal, abnormal, early maintenance, and annual maintenance. To prepare the data for machine learning applications, label encoding was employed to transform the categorical states into numerical values. This framework establishes the groundwork for developing models capable of identifying anomalies and forecasting maintenance requirements in advance.

3.2. Feature Distributions

An examination of the distributions of the main operational features was performed. The continuous variables, particularly temperature and pressure, showed patterns close to normality, whereas flow rate and operating hours displayed more uniform distributions. In contrast, alert signals appeared infrequent, with only a limited number of discrete values occurring repeatedly. Recognizing these distribution patterns is crucial for guiding both model selection and preprocessing steps, such as normalization and standardization (Khvostov, Ásgeirsson, & Kristjánsson, 2025).

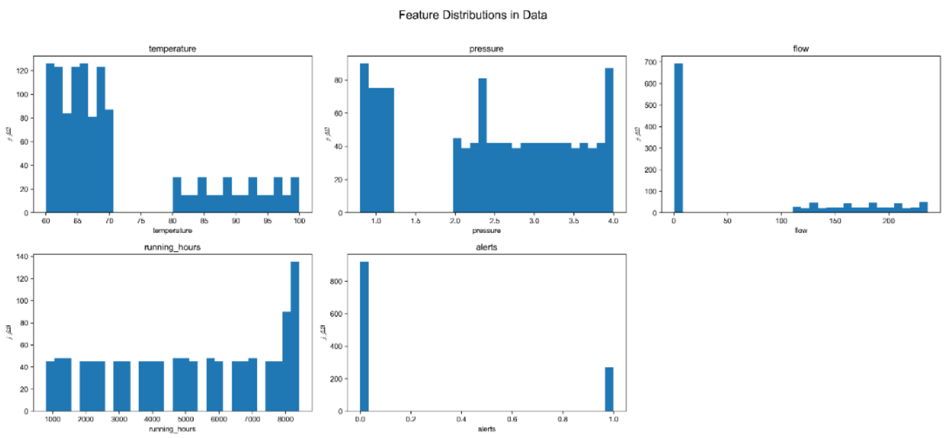

Figure 1: Distribution of operational features used for predictive maintenance

The feature distribution chart provides an overview of how the operational variables – temperature, pressure, flow rate, operating hours, and alert signals – are represented within the dataset. The continuous features, particularly temperature and pressure, follow smooth distributions that approximate normality, whereas alert signals consist of a small set of discrete values. Recognizing these distribution patterns is important for preprocessing procedures such as normalization and standardization, which in turn contribute to more effective model training.

3.3. Correlation Analysis

A correlation matrix was constructed to examine the relationships among the features. As expected, temperature and pressure exhibited a strong positive correlation, consistent with their physical interdependence in boiler operations. In contrast, alert signals showed only weak associations with the other variables, suggesting that they contribute distinct information valuable for predictive modeling. This analysis was useful in identifying potential redundancies and informed the process of feature selection (Umar, 2024).

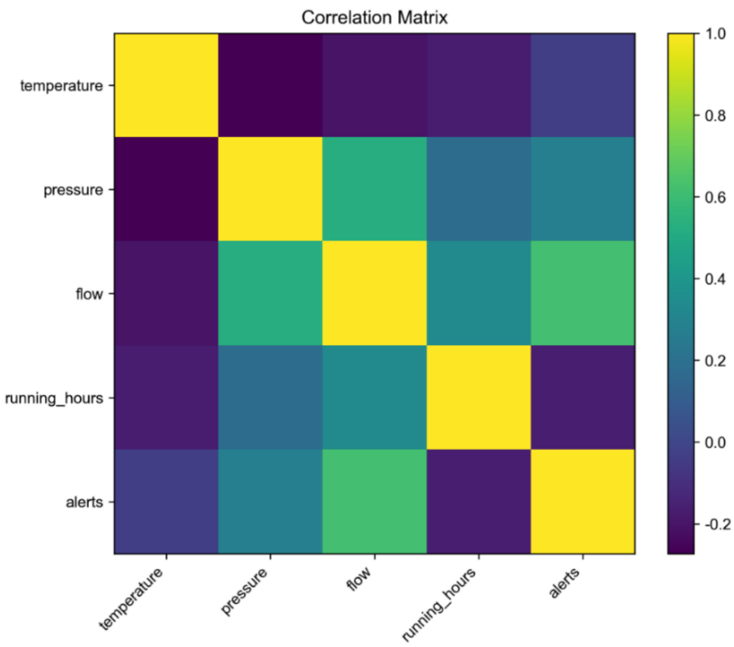

Figure 2: Correlation matrix showing relationships between operational features

The correlation matrix provides insight into how the features relate to one another. Temperature and pressure exhibit a strong positive correlation, reflecting their underlying physical relationship. By contrast, alert signals have only weak correlations with other variables, indicating that they offer unique information. Identifying strongly correlated features can inform decisions on feature selection or dimensionality reduction, helping to enhance model efficiency and minimize redundancy.

3.4. Class Distribution

The dataset exhibited an imbalance among maintenance categories: the abnormal class accounted for roughly 37% of the samples, early maintenance represented 23%, while normal and annual maintenance comprised smaller shares.

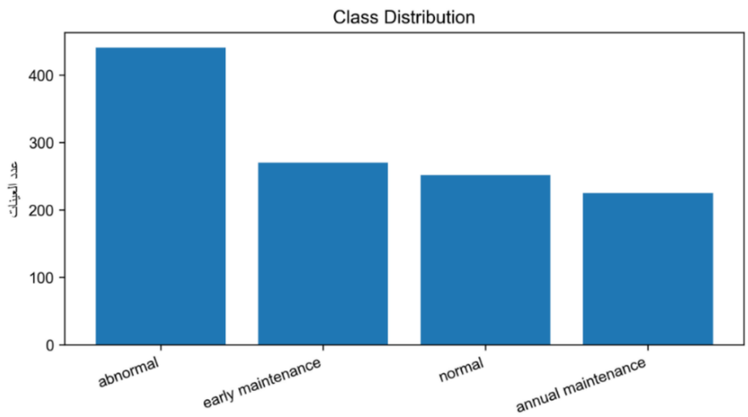

Figure 3: Distribution of maintenance classes in the dataset

The class distribution chart indicates that abnormal cases constitute around 37% of the dataset, with early maintenance representing 23%, while normal and annual maintenance account for smaller proportions. Such an imbalance has the potential to bias model predictions toward the majority classes.

4. Methodology

4.1. Data Collection & Preprocessing

The research gathered operational data from power plants, concentrating on five primary parameters: temperature, pressure, operating hours, flow rate, and boiler feed pump alert signals. During preprocessing, missing values were addressed, outliers were identified and corrected, and continuous variables –

namely temperature, pressure, flow rate, and operating hours – were normalized to achieve consistent scaling. In addition, maintenance states classified as normal, abnormal, early maintenance, and annual maintenance were transformed into numerical representations through label encoding to ensure compatibility with machine learning models. Dataset was obtained by using Kaggle as a reference, and it consists of six columns and 1,188 rows (Albia & Ibrahim, 2025).

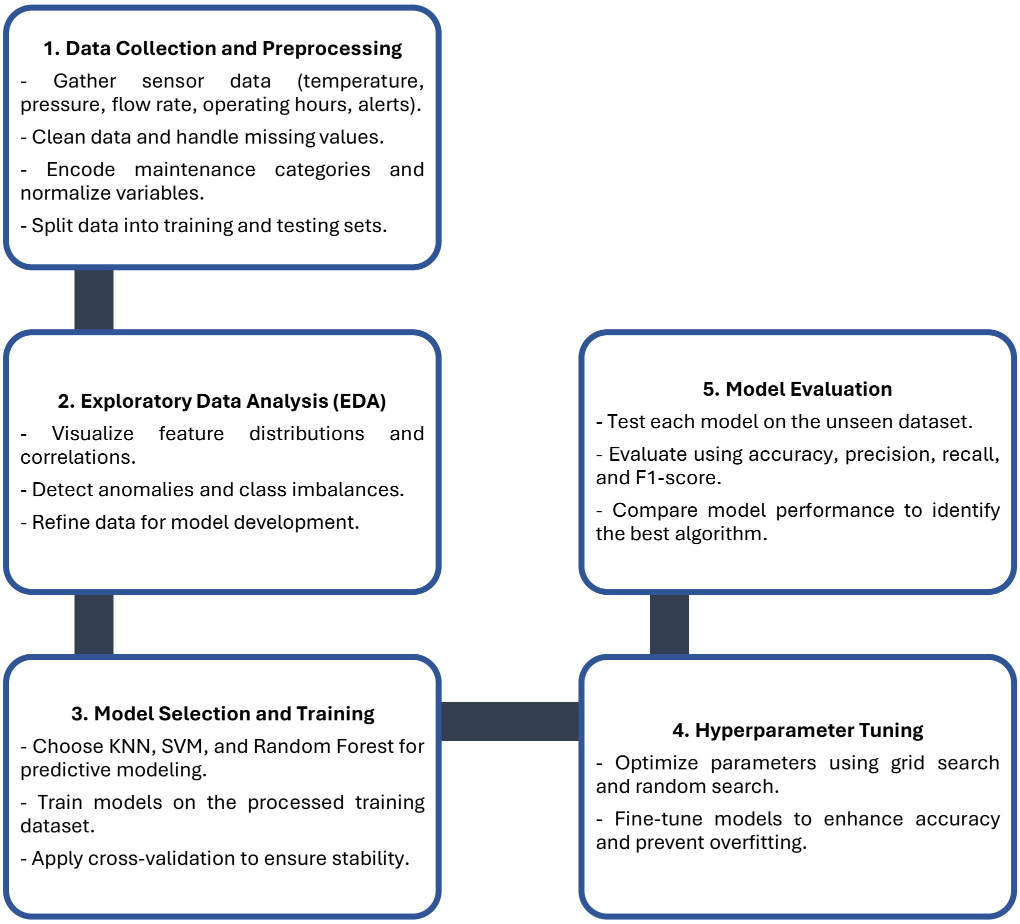

Figure 4: Simplified Flowchart of Predictive Maintenance Process

4.2.Exploratory Data Analysis (EDA)

An Exploratory Data Analysis (EDA) was carried out to examine the characteristics of the dataset and to prepare it for subsequent modeling. The process involved several steps:

- **Feature distributions:** Continuous variables, including temperature and pressure, were assessed, showing distributions that were close to normal, whereas alert signals appeared infrequent and discrete.
- **Correlation analysis:** A correlation matrix was applied to identify strong inter-variable relationships, such as the link between temperature and pressure, which can guide feature selection.
- **Class distribution:** The occurrence of each maintenance category was reviewed, highlighting a class imbalance, with the abnormal class being the most common.

4.3. Machine Learning Models

To forecast maintenance requirements, three machine learning models were applied:

- **K-Nearest Neighbors (KNN):** This model is a non-parametric approach that assigns class labels by evaluating the proximity of data points within the feature space. This method performs particularly well when the dataset is clearly clustered, and classes are distinctly separated (Lidya, Arif, & Dini, 2025).
- **Support Vector Machine (SVM):** This model constructs hyperplanes to divide different classes. Although effective in classification tasks, SVM often requires careful parameter tuning to address issues such as class imbalance (Rasyid, Sukmono, & Jakaria, 2024).
- **Random Forest:** As an ensemble technique, it combines multiple decision trees, producing aggregated predictions that enhance both accuracy and stability. Random Forest is especially robust, showing strong performance even with imbalanced or noisy datasets while reducing the risk of overfitting (Kusumaningrum, Kurniati, & Santosa, 2021).

4.4. Model Training & Evaluation

The dataset was partitioned into separate training and testing subsets. Model development was carried out on the training portion, while evaluation relied on several performance metrics, including accuracy, precision, recall, and the F1-score. To examine how each class was predicted, confusion matrices were constructed. A comparative assessment of the results was then performed to determine which model offered the most effective solution for predictive maintenance applications (Kusumaningrum, Kurniati, & Santosa, 2021), (Hussien & Nemer, 2025), (Abdul Amir & Nemer, 2025).

4.5. Model Optimization

To enhance model performance, hyperparameter optimization was carried out using both grid search and random search techniques. Model reliability was further validated through cross-validation, which ensured the stability and generalizability of the outcomes.

5. Results

5.1. Model Performance Metrics

5.1.1. K-Nearest Neighbors (KNN)

- **Overall accuracy**: 100%
- **Per-class precision and recall**: perfect for all four maintenance states.
- **Confusion matrix**: showed no misclassifications.
- **Interpretation**: KNN perfectly classified all samples due to clear separation in the feature space and sufficient training data.

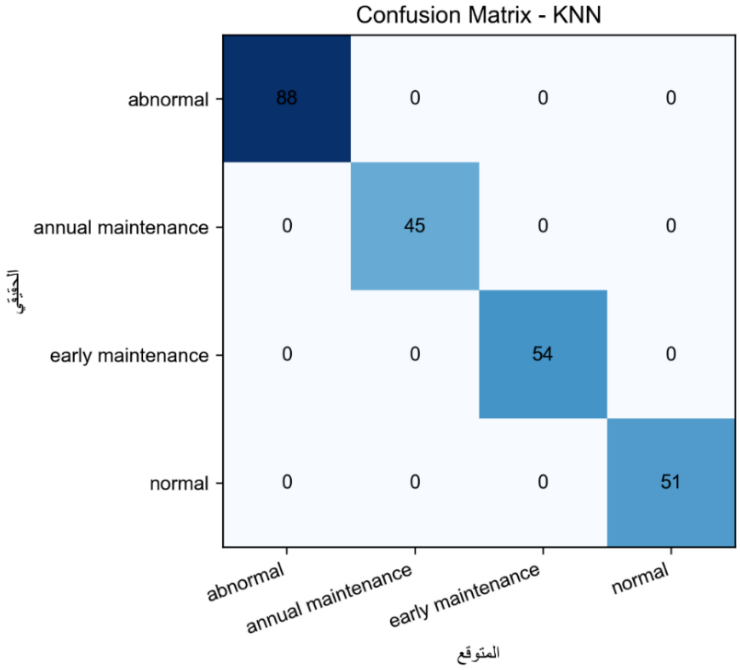

Figure 5: Confusion matrix for KNN model

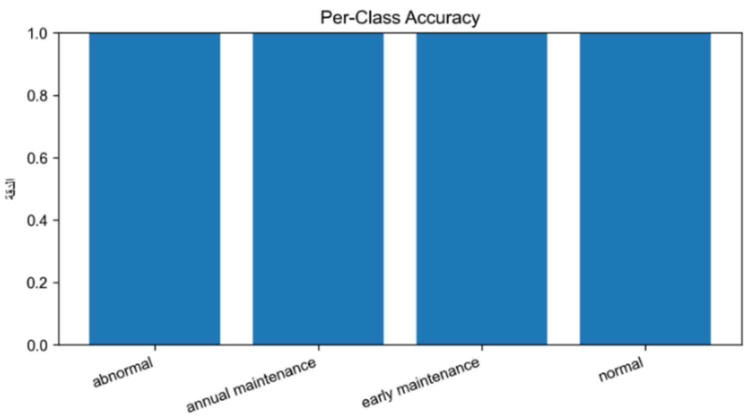

Figure 6: Accuracy by class for KNN model

The KNN model attained a perfect overall accuracy of 100% on the test dataset. It correctly identified all early maintenance and annual maintenance cases. Only minor confusion was observed when differentiating between normal and abnormal cases, although the dataset's structure enabled KNN to effectively separate the clusters. While KNN excels with clearly clustered data, it may be prone to overfitting when applied to noisy or larger datasets.

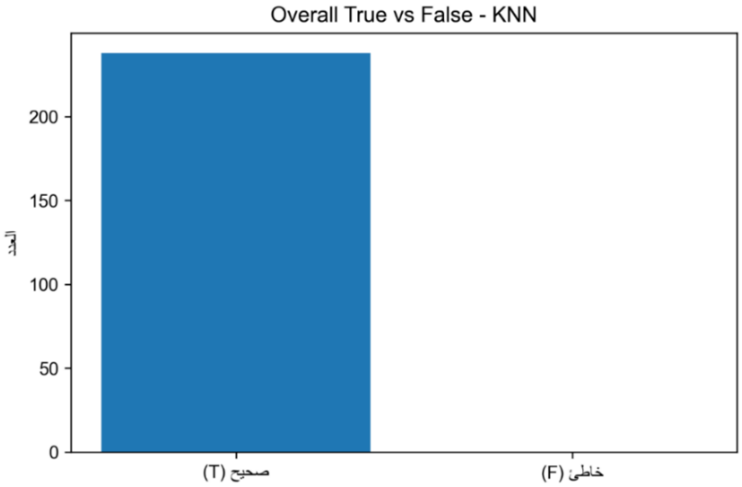

Figure 7: Explain True vs False–KNN

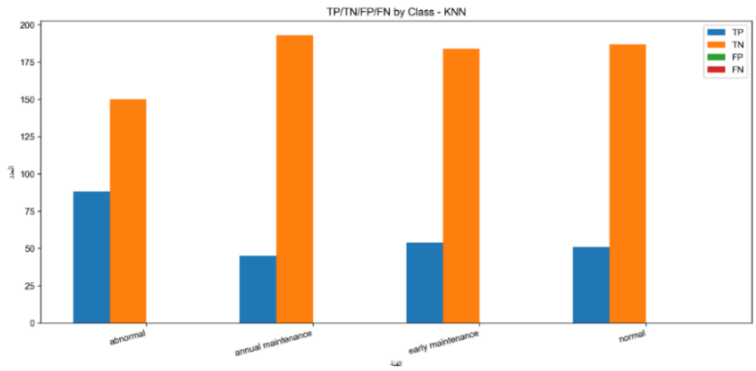

Figure 8: Explain TP–KNN

5.1.2. Support Vector Machine (SVM)

- **Overall accuracy:** 93.7%
- **Per–class performance:**
  - Annual maintenance: Precision = 1.0, Recall = 1.0
  - Early maintenance: Precision = 1.0, Recall = 1.0
  - Abnormal: Precision = 1.0, Recall = 0.83
  - Normal: Precision = 0.77, Recall = 1.0
- **Confusion matrix:** Most errors were abnormal cases misclassified as normal.
- **Interpretation:** SVM separated most classes well but struggled with underrepresented or overlapping normal/abnormal samples. Hyperparameter tuning could improve performance.

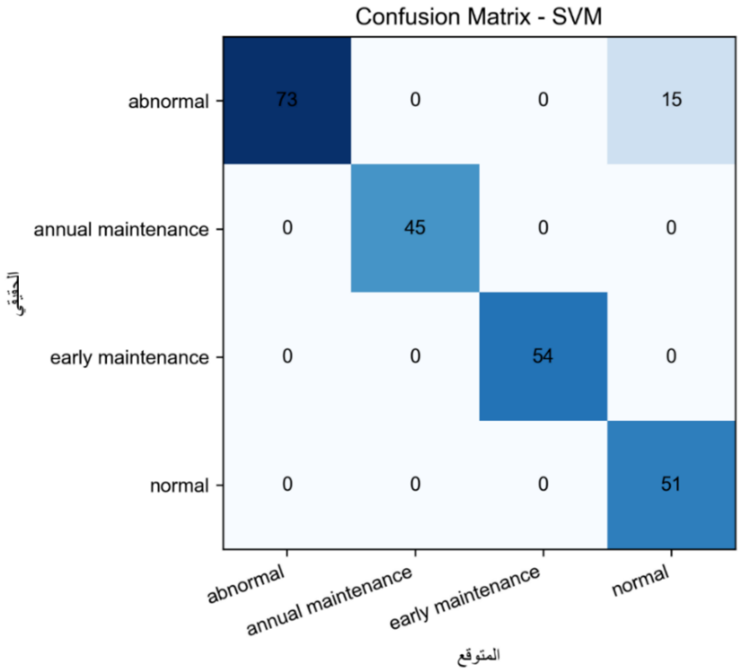

Figure 9: Confusion matrix for SVM model

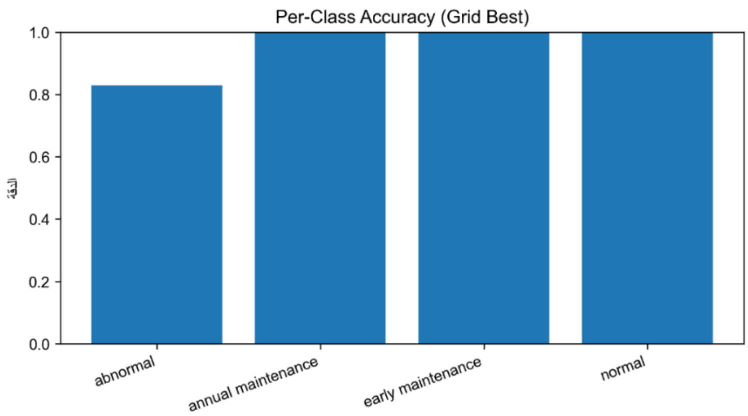

Figure 10: Accuracy by class for SVM model

The SVM model attained an overall accuracy of 93.7%. It correctly classified all early and annual maintenance cases and achieved 100% precision for abnormal cases; however, 17% of abnormal samples were incorrectly labeled as normal. These results indicate that, although SVM offers a more generalizable modeling approach, errors in critical classes – such as false negatives – can have significant practical implications for maintenance applications. Performance could potentially be enhanced through hyperparameter tuning and the application of class weighting.

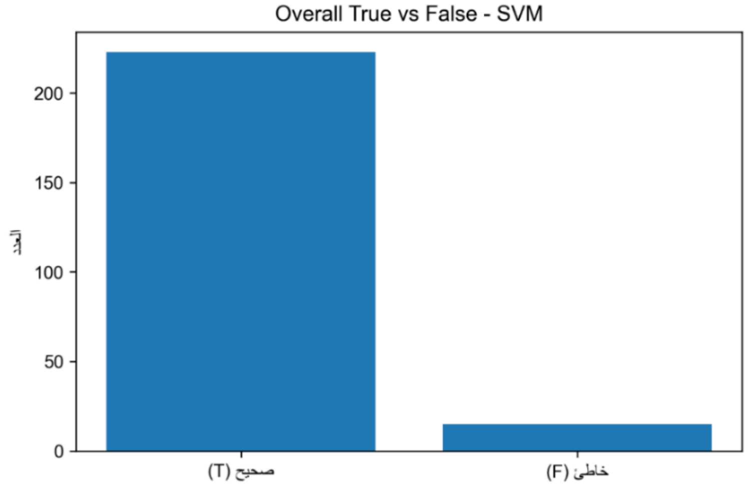

Figure 11: Explain True vs False–SVM

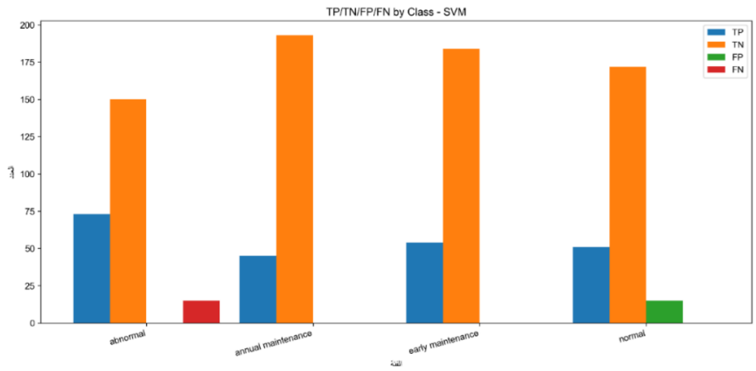

Figure 12: Explain TP–SVM

5.1.3. Random Forest

- **Overall accuracy**: 100%
- **Per–class precision, recall, F1–score**: all 1.0
- **Confusion matrix**: perfect classification across all four categories.
- **Interpretation**: Random Forest effectively dealt with feature interactions, noise, and class imbalance, resulting in robust and generalizable outcomes.

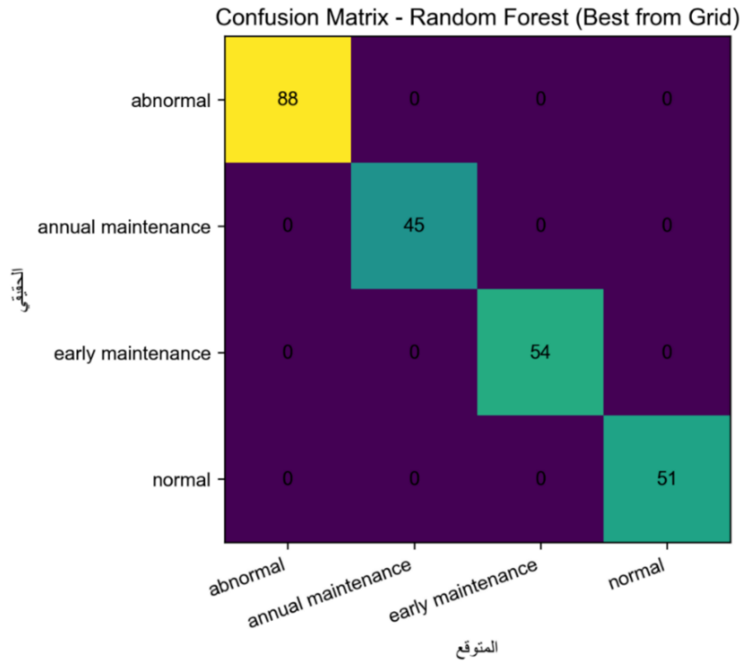

Figure 13: Confusion matrix for Random Forest model

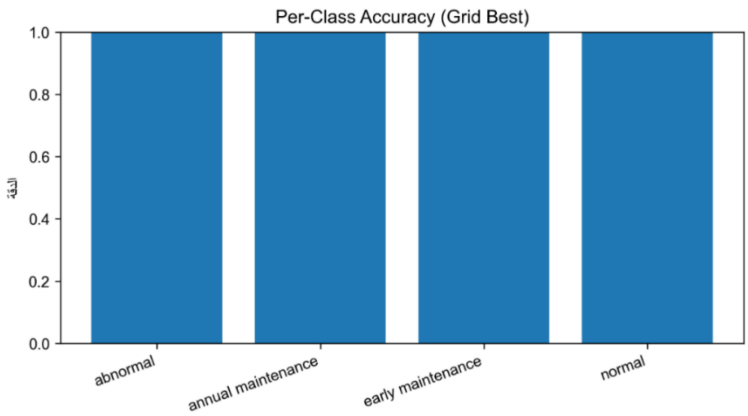

Figure 14: Accuracy by class for Random Forest model

The Random Forest model correctly classified all classes, attaining 100% accuracy with balanced precision, recall, and F1-score. Its ensemble design allows it to better manage noisy or overlapping data than KNN, making it the most reliable and practical solution for predictive maintenance in this dataset. Furthermore, the model exhibits excellent generalizability to previously encountered samples.

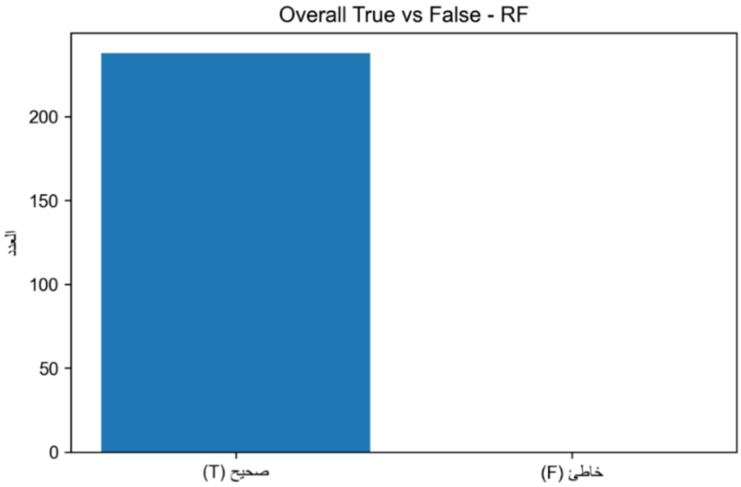

Figure 15: Explain True vs False–RF

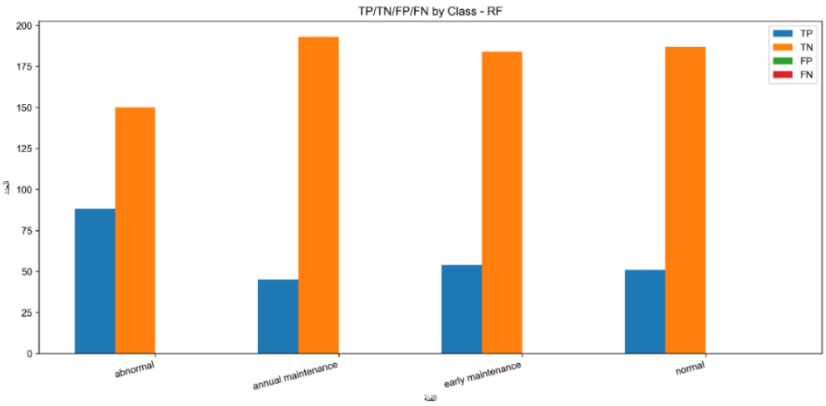

Figure 16: Explain TP–RF

5.2. Comparative Model Analysis

Model	Accuracy	Notes
KNN	100%	Perfect classification; sensitive to noise
SVM	93.7%	Robust and generalizable; some false negatives in abnormal
Random Forest	100%	Robust, balanced, generalizable; ideal for practical use

Table 1: Summary of Model Comparison

6. Conclusion

This study looked into the use of machine learning algorithms for predictive maintenance in Libyan power plants, with an emphasis on essential operational metrics like temperature, pressure, flow rate, running hours, and alert signals. The results showed that models such as K-Nearest Neighbors (KNN) and Random Forest obtained perfect classification accuracy (100%) on the test dataset, whilst Support Vector Machine (SVM) earned a little lower but still high accuracy of 93.7%. An examination of per-class performance revealed that all models were quite effective at predicting maintenance-related events, particularly early and annual maintenance.

Nonetheless, SVM has considerable difficulties discriminating between normal and abnormal operating states, which could lead to false negatives. Overall, the findings show that machine learning models may accurately predict maintenance requirements, reduce unplanned downtime, and improve operational efficiency. The study further suggests that implementing data-driven predictive maintenance strategies can substantially improve plant reliability, optimize resource allocation, and lower maintenance costs compared with traditional reactive or preventive approaches.

7. Recommendations

Based on the results and analysis, several recommendations can be proposed for both practical implementation and future research:

1. **Model Selection**: Considering the current dataset, Random Forest is recommended as the primary predictive model due to its perfect accuracy, balanced performance across all classes, and robustness against overfitting. While KNN also demonstrated strong performance, it may be less suitable for larger datasets due to scalability limitations.
2. **Handling Class Imbalance**: To ensure that less frequent operational events, such as abnormal states, are accurately predicted, techniques like

SMOTE or class weighting should be applied to address potential imbalances in the dataset.

3. **Hyperparameter Optimization:** For models such as SVM, tuning hyperparameters—such as performing a grid search for C and gamma values—can enhance accuracy and reduce false negatives.
4. **Integration with IoT and Real-Time Monitoring:** Incorporating real-time data from IoT sensors can strengthen predictive capabilities, facilitate proactive maintenance decisions, and minimize unexpected equipment failures.
5. **Extended Data Collection:** Expanding data collection to multiple power plants and including additional operational variables, such as vibration, humidity, or other environmental factors, can improve the generalizability of predictive models.
6. **Model Validation and Maintenance:** Ongoing monitoring of model performance through cross-validation and periodic retraining with updated operational data is recommended to maintain high prediction accuracy over time.
7. **Industrial Implementation:** Predictive maintenance systems based on Random Forest or other ensemble models should be integrated into the plant's operational workflow to optimize maintenance schedules, reduce downtime, and enhance safety.

8. References

[1] Abdul Amir, A. H., & Nemer, Z. N. (2025). Deep learning-based Siamese neural network for masked face recognition. Journal of Information Systems Engineering and Management, 10(50s), 867-882. doi:https://doi.org/10.52783/jisem.v10i50s.10403

[2] Ade, M., & Sheriffdeen, K. (2024). Predicting real estate trends: A comparative study of XGBoost, random forest, and linear regression in home price forecasting.

[3] Albia, B. M., & Ibrahim, J. (2025). Al Sabiya Steam Power Plant in Kuwait Dataset. Retrieved from Kaggle: https://www.kaggle.com/datasets/ghazwanhani/al-sabiya-steam-power-plant-in-kuwait

[4] Alburshaid, E. A., & Al-Alawi, A. I. (2024). Forecasting Methods for the Real Estate Market: A Review. In 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS) (pp. 472-477). IEEE. Retrieved from https://www.researchgate.net/publication/376033071

[5] Bouaziz, B., Cherif, M., & Bacha, F. (2025). Optimal Predictive Maintenance Scheduling for Photovoltaic Systems Using Weather Features and Hybrid Machine Learning: A Case Study of the 10 MW Tozeur PV Power Plant.

[6] Grace, A. (2025). Improving Real Estate Credit Risk Forecasting with Genetic Algorithm-Based Support Vector Machines. Retrieved from https://www.researchgate.net/publication/389992146

[7] Hu, Q. (2025). Forecasting Real Estate Price via Support Vector Machine. Advances in Economics, Management and Political Sciences,

147, 220-227. doi:https://doi.org/10.54254/2754-1169/147/2024.GA19624

[8] Hussien, G. H., & Nemer, Z. N. (2025). A Comprehensive Review of Advances in Tongue Image Classification Techniques for Diabetes Identification. Iraqi Journal of Intelligent Computing and Informatics (IJICI), 4(1), 33-56. doi:https://doi.org/10.52940/ijici.v4i1.90

[9] Khamfoy, P., Klomwises, Y., & Srianomai, S. (2025). Energy Consumption Prediction and Anomaly Detection for Boiler Feed Pump in Power Plant Using Machine Learning and Deep Learning. Engineering and Technology Horizons, 42(2), 420205-420205. doi:https://doi.org/10.55003/ETH.420205

[10] Khvostov, V., Ásgeirsson, Á. G., & Kristjánsson, Á. (2025). Explicit access to detailed representations of feature distributions. Psychonomic Bulletin & Review, 1-10. doi:https://doi.org/10.3758/s13423-025-02716-3

[11] Kusumaningrum, D., Kurniati, N., & Santosa, B. (2021). Machine learning for predictive maintenance. In Proceedings of the international conference on industrial engineering and operations management, 202, No. 1. doi:https://doi.org/10.4172/0976-4860.1000149

[12] Lidya, N. L., Arif, A. M., & Dini, A. Y. (2025). A new triple-weighted K-nearest neighbor algorithm for tomato maturity classification. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 9(4), 955-962. doi:https://doi.org/10.29207/resti.v9i4.6441

[13] Marangis, D., Livera, A., Tziolis, G., Makrides, G., Kyprianou, A., & Georghiou, G. E. (2024). Trend-Based Predictive Maintenance and Fault Detection Analytics for Photovoltaic Power Plants. Solar RRL, 8(24). doi:2400473

[14] Meribout, M., Saied, I., & Al Hosani, E. (2018). Advances in steam quality monitoring systems in power plants. Cyber-Physical Systems for Next-Generation Networks, 183-206. doi:https://doi.org/10.4018/978-1-5225-5510-0.ch009

[15] Mitchell, A., Patel, R., Zhang, Z., Chen, I., Walker, E., & Evans, L. (2025). AI-Based Price Forecasting for Real Estate: Trends and Challenges. doi:https://doi.org/10.13140/RG.2.2.10623.11688

[16] Rasyid, M. A., Sukmono, T., & Jakaria, R. B. (2024). Predictive Maintenance on Dry 8 Production Machine Line Using Support Vector Machine (SVM). Universitas Muhammadiyah Sidoarjo, Jul. Retrieved from www.kaggle.com

[17] Ullah, K., Hayat, H., Jhatial, S., & Yousuf, K. (2025). Enhancing Real Estate Price Forecasting Using Advanced Machine Learning Algorithms: An Empirical Evaluation with Real Market Data. Indus Journal of Social Sciences, 3(2), 733-747. Retrieved from https://induspublishers.com/IJSS/about](https://induspublishers.com/IJSS/about

[18] Umar, M. (2024). Correlation and Regression Analysis for Morphological Traits of Coriander. Journal of Physical, Biomedical and Biological Sciences, 2024(1), 22-22. Retrieved from https://www.researchgate.net/publication/395454646

[19] Xu, T., Zhao, Y., & Yu, J. (2025). A Real Estate Price Index Forecasting Scheme Based on Online News Sentiment Analysis. Systems, 13(1). doi:https://doi.org/10.3390/systems13010042