IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

Predictive Maintenance for Industrial Equipment Using Machine Learning

Dhruv Patel[1] and Prathmesh Kalgutkar[2]
Department of Computer Science and Engineering (Cyber Security)
Thakur College of Engineering and Technology, Mumbai, India[1]
Department of Artificial Intelligence and Data Science
New Horizon Institute of Technology And Management, Mumbai, India[2]
pateldhruv0609@gmail.com and prathmesh161203@gmail.com

Abstract*: By combining condition monitoring with predictive maintenance, the industry may significantly increase system dependability and prevent significant financial losses from unplanned motor failures. The industry uses electric engines and additional machinery. This paper presents a Random Forest-based anticipatory maintenance using machine learning architecture. Through the development of data gathering and analysis procedures, use of machine learning methodology, and comparison with simulation tool analysis, the system was evaluated on an actual industrial example. Data Analysis Tool is able to access data that has been gathered by a variety of sensors, machine PLCs, and communication protocols on the Azure Cloud architecture. Initial findings demonstrate that the method behaves appropriately when forecasting various machine states with a high degree of accuracy.*

Keywords: predictive maintenance, machine learning, random forest

I. INTRODUCTION

Predictive maintenance (PdM), commonly referred to as "on-line monitoring," "risk-based maintenance," or "condition-based maintenance," has a long history and has been the subject of several recent research projects. It discusses maintaining careful watch on machinery to avoid future problems. Predictive maintenance began with the initial method of visual inspection and has evolved into automated systems that use advanced signal processing techniques based on neural networks, fuzzy logic, machine learning, and pattern recognition. Automated technologies provide a practical solution for many industries when human eyes and ears are no longer able to detect and acquire sensitive information from equipment, notably motors. When automated sensors and automated repair are used together, it can decrease machine downtime, avoid replacing equipment that isn't necessary, locate the problem's source, save costs, and boost productivity. Predictive maintenance and preventative maintenance overlap in terms of scheduling the repair task in advance to avoid machine issues. Unlike typical preventative maintenance, predictive maintenance solutions rely on sensor data and analysis algorithms. In the manufacturing industries, induction engines account for over 70 percent of all driven electrical consumption. In this context, there seems to be a lot of interest in techniques to more precisely determine the condition of health of these motors. It is discovered that breakdown of bearings is the most prevalent cause of motor malfunction and the most common maintenance problem. Therefore, the two key areas of focus for predictive maintenance are minimizing unscheduled downtime and increasing energy efficiency, which is crucial for energy saving.Furthermore, the algorithms developed in relation to these two can be broadly divided into two groups:
1) energy and efficiency: a variety of methods and instruments for assessment have been created;
2) System status monitoring, which comprises established fault-detection approaches for motor fault detection.
Algorithms are required for predictive maintenance to function properly. As mentioned, distinct strategies, such as data processing, diagnostics, and prognostics, ought to be applied at various phases of PdM deployment. In PdM, three different strategies can be distinguished. There are three methods available:
1) Data-driven;
2) Model-based
3) Hybrid.

IJARSCT ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53 Volume 4, Issue 1, August 2024

The data-driven strategy, commonly known as the machine learning approach or the data mining technique, uses past data to build a model of system behaviour. An understanding of the target product physically can be a part of a model-based approach, which represents the behaviour of the system using an analytical model. Machine learning approaches are useful in fields where data availability is increasing, such as maintenance. It provides newly created algorithms, cloud-based solutions, and ever-more-efficient solutions.The following are the two main categories of PdM based on machine learning: Where process and/or logistical information is available but no maintenance data is available, the situation is referred to as unsupervised. Supervised failure occurrence information is included in the modelling dataset. The features of the existing maintenance management policy mostly dictate how easily accessible maintenance information is. If at all possible, supervised solutions are recommended. As far as machine learning is concerned, depending on the output of the data set, two kinds of supervised issues are possible: regression problems (if the output assumes continuous values) and classification problems (if the output assumes categorical values).

II. LITERATURE SURVEY

The following is a quick summary of some studies that have been published in the past on the predictive maintenance of industrial machinery by different researchers using different approaches.

Regarding a semiconductor production implant maintenance task, Gian Antonio Susto has described a system that has been shown to outperform both conventional PvM techniques and a distance-based PdM alternative utilizing just one Support Vector Machine classifier [1]. The present case research additionally demonstrated that SVMs outperform k-NN classifier when executing MCPdM, and that MC-PdM-known regularly outperforms PvM approaches on average.

In addition to the data-type components needed for predictive modelling, the Industrial Internet of Things is leveraged to extract useful insights from machine information [2]. Autoregressive integrated moving average (ARIMA) forecasting trauma plating machines have been investigated by the researchers as a potential tool for quality defect, maintenance, and downtime prediction. It has been demonstrated that machine learning is a crucial part of the commercial internet of things for quality control and management. Both performance and the manufacturing process are improved by it.

In an alternate investigation by Marina Paolanti, precise estimations were obtained by applying the concept in an experimental setting with a real industrial group as an example [3]. A variety of sensors, machine PLCs, and communication protocols have all gathered data, which has been made available to the Data Analysis Tool. By training a Random Forest technique on Azure Machine Learning Studio, the suggested PdM methodology enables the adoption of dynamic decision rules for maintenance management. Initial findings demonstrate the approach's correct behaviour in predicting various machine states with a high degree of accuracy (95%) on a set of 530731 data readings on 15 distinct machine features that were obtained in real time from the cutting machine under test.

In his research, K. Liulys demonstrates that routine equipment maintenance is necessary to prevent machine failure. Preventive maintenance, however, necessitates a lot of pointless and repetitive checking [4]. It frequently results in increased expenses that are unaffordable. As a result, this equipment can benefit from the novel idea of predictive maintenance, which lowers the cost of downtime and the frequency of checks. The amount of time and money needed to do all of the checking is significantly decreased with predictive maintenance. The method of predicting output values by running algorithms on massive amounts of collected core value data is known as machine learning.

A cognitive analytics-based system for keeping track of machines and recognizing anomalies has been introduced by Farzam Farbiz. We verified the suggested methodology by using the suggested framework on a robot for industrial use case [5]. The machine model created by the suggested framework has the ability to detect anomalies in the robot's movement in real time and learn to change its performance in response to fresh data. It should be emphasized that the suggested technique has only been used on one use-case research, albeit showing encouraging findings and an unsupervised machine learning model that can categorize the data in spite of noise and outliers.

III. IMPLEMENTATION

This research was put into practice by gathering and preprocessing data, choosing machine learning models, training them, and then assessing how well they worked for predictive maintenance on industrial machinery.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

3.1 Proposed Architecture

Data Acquisition Layer

- Sensor Networks: Collects real-time sensor data from industrial equipment, including temperature, vibration, pressure, and other relevant parameters.
- IoT Devices: Integrates with IoT devices for continuous monitoring of equipment health and performance.
- SCADA Systems: Interfaces with Supervisory Control and Data Acquisition (SCADA) systems to access operational data and control signals.

Data Preprocessing and Storage

- Data Cleansing: Cleans and preprocesses raw sensor data to remove noise, outliers, and inconsistencies.
- Feature Engineering: Extracts relevant features from the preprocessed data, including statistical features, time-domain features, and frequency-domain features.
- Data Storage: Stores preprocessed data in a scalable and fault-tolerant data repository, such as a data lake or time-series database, for efficient storage and retrieval.

Machine Learning Model Development

- Model Selection: Selects appropriate machine learning algorithms based on the nature of the predictive maintenance task, such as classification, regression, or anomaly detection.
- Training Data Preparation: Prepares labeled training data by associating historical sensor data with maintenance events, failures, or operational states.
- Model Training: Trains machine learning models using the prepared training data to learn patterns and relationships between sensor data and equipment health.

Predictive Analytics and Prognostics

- Failure Prediction: Utilizes trained machine learning models to predict impending equipment failures or degradation.
- Anomaly Detection: Detects abnormal equipment behavior or deviations from normal operating conditions using anomaly detection techniques.

Decision Support and Maintenance Planning

- Maintenance Planning: Generates optimized maintenance schedules and resource allocation strategies based on predicted failure probabilities, operational priorities, and available resources.
- Decision Support: Provides actionable insights and recommendations to maintenance personnel and operators through intuitive dashboards and visualization tools.

Integration and Deployment

- Integration with Operational Systems: Integrates predictive maintenance system with existing operational systems, such as enterprise asset management (EAM) systems or maintenance management systems (MMS).
- Deployment: Deploys trained machine learning models and predictive analytics algorithms in production environments, ensuring scalability, reliability, and real-time performance.

Monitoring and Continuous Improvement

- Model Monitoring: Monitors the performance of deployed machine learning models in production, detecting drifts or degradation in model accuracy over time.
- Feedback Loop: Collects feedback from maintenance activities and operational data to continuously refine and improve predictive maintenance models.
- Model Retraining: Periodically retrains machine learning models with updated data to adapt to changing operational conditions and improve predictive accuracy.

Security and Compliance

- Data Security: Implements robust security measures to protect sensitive equipment data from unauthorized access, manipulation, or cyber threats.
- Compliance: Ensures compliance with industry regulations and standards governing data privacy, security, and operational safety in industrial environments.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

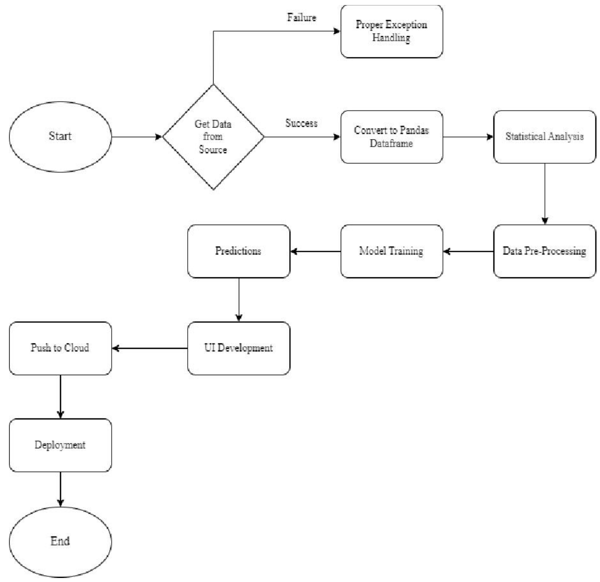

Figure 1. Workflow of Predictive Maintenance System for Industrial Equipment

3.2 Algorithms Used

Logistics Regression

The statistical method of logistic regression can be utilized for analyzing a dataset in cases where the outcome is influenced by one or more independent factors. In machine learning and data analytics, it is often used for classification problems when the outcome is a binary (yes or no) or categorical variable.The logistic regression model determines the likelihood of the outcome variable based on the independent components. It is a type of generalized linear model that uses the logistic function to translate the linear output of the independent components into a probability value between 0 and 1.

The logistic regression equation can be obtained from the linear regression equation. The mathematical techniques to derive equations for logistic regression are as follows:

The straight-line equation is expressed as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + \ \dots\dots\dots\dots\dots + \ b_nx_n$$

Since y in a logistic regression can only be between 0 and 1, we can divide the previous equation by (1-y) to account for this:

$$\frac{y}{1-y}; 0 \ for \ y = 0 \ and \ infinity \ for \ y = 1$$

However, we require a range from -[infinity] to +[infinity]. After taking the equation's logarithm, we get the following:

$$\log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + \ \dots\dots\dots\dots\dots + \ b_nx_n$$

The final equation for logistic regression is mentioned above.

Decision Tree

When performing predictive maintenance on industrial machinery, decision trees can be utilized to spot any problems before they arise. Machine learning techniques such as decision tree models can be applied to regression and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

classification problems. In order to get the final forecast, they divide the input data into a series of binary decisions, each of which divides the data into smaller subsets, until they reach a terminal node.

Decision tree models can be trained using past machine performance and maintenance records. This allows decision trees to be utilized in predictive maintenance. On the basis of real-time data gathered from the machine's sensors, the model is then able to forecast when maintenance is necessary.

Selecting the ideal attribute for the root node and sub-nodes of a decision tree is the main challenge throughout its implementation.

To solve these problems, an approach called attribute selection measure, or ASM, might be applied. We can quickly select the best feature for the tree nodes by using this measurement. The two most popular ASM techniques are as follows:

a) Information Gain: Information gain calculates changes in entropy following an attribute-based dataset segmentation. It calculates the amount of information a characteristic provides about a class. Entropy is a measure of the impurity of a given property. Unpredictability of data is defined there. To calculate entropy, use

$$H = -\sum_{t=1}^{n} p(x_t)\log_2 p(x_t)$$

*b) Gini Index:*The Gini index serves as a purity or impurity indication when a decision tree is created using the CART (Classification and Regression Tree) technique. Having a trait with a low Gini index is better than having one with a high value.

$$Gini = 1 - \sum_{t=1}^{n} p_t^2$$

3.3 Random Forest

Random Forest is a machine learning technique that is widely used for predictive maintenance. This is an illustration of an ensemble learning technique that boosts prediction accuracy by combining several decision trees.

Random Forest can be used in predictive maintenance to anticipate machine faults or failures based on historical data. Large volumes of data from sensors, devices, and other sources are analysed by the algorithm to look for patterns and trends that could indicate problems. With this information, the system then predicts possible malfunctions or breakdowns.

The capability of Random Forest to handle big, complicated datasets with lots of variables is one of its main benefits when it comes to predictive maintenance. Both numerical and category data may be handled by the algorithm, which can also automatically identify the key characteristics for producing predictions.

You must perform the following in order to use Random Forest for predictive maintenance:

- Gather and clean up past data on equipment malfunctions and maintenance tasks.
- Determine pertinent characteristics and circumstances that could indicate future failures.
- Use the historical data to train the Random Forest model.
- Run the model's performance and accuracy tests using fresh data.
- Make predictions about upcoming malfunctions or breakdowns using the model.

3.4 Support Vector Machine

The SVM model is commonly employed in binary classification tasks. Industrial equipment has made extensive use of SVMs to determine a specific state based on the received signal. Because the SVM model can map low-dimension features to hyperplanes and handle a range of fault kinds, it can also be utilized to finish multiclass jobs. In summary, the primary objective of support vector machines (SVM) is to identify a hyperplane and divide data points suitably on both of its sides. The optimization object is represented as

$$argmax\ (w,b)\left\{\frac{1}{\|w\|} min\left[y_{t(w^T\ .\ x_t+b)}\right]\right\}$$

$$s\,.\,t\ \ y_t(w^T\ .\ x_t + b)\ \ \geq 1$$

where (x_t, y_t) refers to a sample that contains features and labels.
SVMs function by determining which hyperplane best divides the data into distinct classes. The goal of the algorithm is to maximize the margin that separates the nearest data points from each class from the hyperplane. Additionally, SVMs make use of a kernel function that converts the data into a higher-dimensional space where data point separation might be simpler.

IV. RESULT AND DISCUSSION

We have successfully constructed a prototype prediction system that can transmit data from the Thingspeak cloud to the edge device.
The state of the VMC (Vertical Machining Center) machine, including air temperature, rotational speed, torque, tool wear, and process temperature, was tracked using a predictive maintenance dataset from Kaggle.
The model is trained using several machine learning methods on a specific collection of data. The sorts of VMC machine failures that could happen are listed below:
Tool wear failure (TWF): At a randomly chosen tool wear time of 200–240 minutes, the tool will either be changed or fail.
Heat dissipation failure (HDF): If the tool's rotational speed is less than 1380 rpm and the difference between the air and process temperatures is less than 8.6 K, heat dissipation leads to a process failure.
Power failure (PWF): The amount of power needed for the process is equal to the product of torque and rotational speed (in rad/s). The process fails if this power is either above 9000 W or below 3500 W.
Overstrain failure (OSF): The process fails as a result of overstrain if the product of tool wear and torque is greater than 11,000 minNm for the L product type (12,000 for M, 13,000 for H).
The confusion matrix is used to assess the system's performance.

4.1 Confusion Matrix

A matrix known as the confusion matrix is used to assess how well the classification models perform for a particular set of test data. It is used to depict key predictive parameters such as recall, specificity, accuracy, and precision. Confusion matrices are useful because they offer straightforward comparisons of variables such as True Positives, False Positives, True Negatives, and False Negatives.

Accuracy

The most logical performance metric is accuracy, which is just the ratio of properly predicted observations to all observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

The ratio of accurately anticipated positive observations to all predicted positive observations is known as precision. (TP/TP+FP = Precision)

$$Precision = \frac{TP}{TP + FP}$$

Recall

The ratio of accurately predicted positive observations to all observations made during the actual class is known as recall.

(Recall is equal to TP plus FN.)

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

The weighted average of Precision and Recall is known as the f1-score. As such, this score accounts for both false positives and false negatives.

The product belonging to the L category was the subject of the majority of the predictions. Our primary goal was to determine whether or not the VMC machine would fail and, if it did, what kind of failure might take place.

The Random Forest Algorithm operates on the basis of multiclass classification and attains maximum accuracy, as demonstrated by the analysis presented in Table 1. The logistic regression algorithm yields the most accurate predictions after random forest.Support Vector Machine (SVM) methods have been found to perform poorly in multiclass classification and to produce less accurate results.

Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	98.52	98	99	98
Logistic Regression	98.08	97	98	98
Decision Tree	97.52	98	98	98
Support Vector Machines	97.04	95	97	96

Table 1. Analysis of Algorithm in (%)

V. CONCLUSION

To sum up, the application of machine learning techniques to develop a Predictive Maintenance System for Industrial Equipment represents a paradigm change in favour of proactive maintenance approaches. This technology gives many advantages to industrial processes by utilizing real-time data analytics and sophisticated algorithms.

The system can evaluate enormous volumes of sensor data and previous maintenance records thanks to machine learning algorithms, which include supervised and unsupervised techniques. This allows the system to estimate remaining usable life, predict equipment breakdowns, and identify anomalies in equipment behaviour. Maintenance staff may take early steps to reduce hazards and avoid costly downtime by having important insights into the health and performance of industrial equipment through continuous monitoring and analysis.

In conclusion, the machine learning-based Predictive Maintenance System for Industrial Equipment is a revolutionary approach to maintenance management that will boost equipment reliability, improve operational efficiency, and ultimately result in large cost savings for businesses in the industrial sector.

VI. FUTURE SCOPE

Machine learning-based predictive maintenance systems for industrial equipment have a bright future ahead of them, with improvements anticipated in a number of areas:

IJARSCT ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53 Volume 4, Issue 1, August 2024

Enhanced Accuracy and Reliability

Upcoming advancements will concentrate on optimizing machine learning algorithms to raise predictive maintenance models' precision and dependability. To capture complex patterns and dependencies in equipment data, this entails combining deep learning architectures, ensemble learning approaches, and advanced feature engineering techniques.

Integration of Multi-Modal Data

Predictive maintenance systems will increasingly leverage multi-modal data sources, including audio, image, and video data, in addition to traditional sensor data. This holistic approach enables a more comprehensive understanding of equipment health and performance, leading to more robust predictive models.

Edge Computing and Real-Time Analytics

With the proliferation of edge computing capabilities, predictive maintenance systems will leverage edge devices to perform real-time data processing and analytics. This enables faster decision-making and reduces latency, particularly in time-sensitive industrial environments where immediate action is required to prevent equipment failures.

Automated Maintenance Planning and Optimization

Anticipated predictive maintenance systems will possess self-governing decision-making abilities to autonomously produce optimal maintenance plans and methods for allocating resources. As a result, maintenance staff have less work to do and proactive maintenance planning based on changing operating circumstances is made possible.

Predictive Maintenance as a Service (PMaaS)

The advent of Predictive Maintenance as a Service (PMaaS) models is expected to provide small and medium-sized organizations (SMEs) with more equitable access to sophisticated predictive maintenance capabilities.SMEs can implement predictive maintenance without having to make a sizable upfront investment in technology and knowledge thanks to cloud-based systems that provide pre-trained machine learning models and scalable infrastructure.

REFERENCES

[1] Gian Antonio Susto et al., “Machine Learning for Predictive Maintenance: A Multiple Classifier Approach,” IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812-820, 2015.

[2] Kaorlis Liulys, “Machine Learning Application in Predictive Maintenance,” 2019 Open Conference of Electrical, Electronic and Information Sciences (es- Tream), pp. 1-4, 2019.

[3] W. Zhang, D. Yang and H. Wang, "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey," in IEEE Systems Journal, vol. 13, no. 3, pp. 2213-2227, Sept. 2019, doi: 10.1109/JSYST.2019.2905565.

[4] AmeethKanawaday, and Aditya Sane, “Machine Learning for Predictive Maintenance of Industrial Machines Using IoT Sensor Data,” in 2017 8th IEEE International Conference on Software Engineering and Service Science, pp. 87-90, 2017.

[5] S. Shukla, A. Maheshwari and P. Johri, "Comparative Analysis of Ml Algorithms & Stream Lit Web Application," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 175-180, doi: 10.1109/ICAC3N53548.2021.9725496.