National Higher School of Technology
Department of Industrial Engineering & Maintenance

Comparative Study of Machine Learning Algorithms for Predictive Maintenance in the Industry 4.0 Context

Master's thesis in
Industrial Engineering

AMROUCHE Ikram, LAHOUASSA Malak Sakina,

Advisor:
Prof. BELAYADI Djahida

Co-advisors:

Academic year:
2022-2023

Abstract: **The fourth industrial revolution is characterized by the increased availability of digital technology and connectivity, leading to a fundamental shift in manufacturing processes. Predictive maintenance (PdM) is identified as a critical factor for successful digital transformation in companies, as it can increase machine life cycles, reduce downtime and associated costs, and boost quality and production rates. In the field of PdM, where proactive planning and timely intervention are crucial to preventing breakdowns, machine learning (ML) algorithms play a vital role. This paper provides an examination of Industry 4.0 technologies and explores their potential to enhance smart maintenance through integration. Additionally, our study conducts a comprehensive comparison of the most frequently utilized ML algorithms for PdM, identifying crucial criteria for achieving model success in applying PdM in an industry 4.0 environment.**

Key-words: predictive maintenance, industry 4.0 technologies, machine learning algorithms

Proceedings of M. F. - 2022/2023

Contents

1 Introduction 2
2 Maintenance 4.0 3
3 Data Processing, Sources, and Types for Predictive Maintenance 5
4 Advanced Machine Learning Algorithms for Predictive Maintenance 6
5 State-of-the-Art in Machine Learning Applied to Predictive Maintenance 9
6 Discussion 10
7 Conclusions 13

1. Introduction

The previous industrial revolution emphasized enhancing physical manufacturing processes, utilizing machinery and steam power, implementing assembly lines for mass production, and introducing automation and electronics. On the other hand, the fourth industrial revolution is focused on creating digital models of physical processes in order to obtain a better understanding and insights into their operations [1].

In light of the trend toward smart manufacturing and the advancement of the Internet of Things (IoT), data mining, and Artificial Intelligence (AI), among other things, Predictive maintenance is proposed as a new kind of maintenance paradigm to only perform maintenance when the analytical models predict specific failures or degradations [2]. PdM is founded on the idea that certain machine characteristics can be tracked and the data collected utilized to calculate an estimate of the equipment's remaining useful life of the equipment[3].according to [4] it refers to the intelligent monitoring of machinery to prevent future breakdowns. Maintenance impacts account for 15 to 60 % of the operational expenses of all production [5]. In order to reduce the costs associated with downtime and defective goods, especially in fiercely competitive advanced manufacturing industries, efficient maintenance activity management is becoming crucial [6].It can avoid unnecessary equipment replacement, reduce machine downtime, find the root cause of the fault, and in this way save costs and improve efficiency [4]. The implementation of the IoT , Cloud Computing, and Big Data in factories allows better monitoring of a production system's current state [7]. To enable early detection of possible failures, PdM systems use prediction tools based on historical data, ad hoc defined health indicators, statistical inference methods, and engineering methodologies. According to [1], statistical inference methods, especially those that incorporate machine learning (ML), are the most efficient in creating models for high-dimensional problems. [7] Highlight the importance of evaluating the criticality of equipment in its operational context before implementing PdM. If equipment failure does not impact plant operation, it may be a viable strategy to run it until it fails and then fix it. However, if a piece of equipment is crucial, it may be necessary to have backup hardware to keep the plant running while the problematic unit is repaired. In this way, the equipment's criticality is reduced by eliminating the single point of failure. Another important aspect in the application of ML-based PdM considered by [1] is the nature of the existing maintenance strategy because the availability of maintenance information mostly depends on it: Supervised learning can be easily used for run to failure (R2F) policies as the data for maintenance cycle is present, whereas for preventive maintenance (PvM) policies, the complete maintenance cycle may not be visible since maintenance is done beforehand to avoid any potential failure. Hence, unsupervised learning approaches are used. However, if available, supervised solutions are better since R2F policies are widely used in industries, and suitable datasets are accessible. However, we verified the need for a more specific discussion of industry 4.0 technologies applied to PdM, and a systematic literature review with a more specific discussion of ML algorithm applied for PdM. As our main contribution, we propose a classification of industry 4.0 technologies, divided into three main classes: data collection and analysis technologies, predictive analysis technologies, and equipment repair technologies. Additionally, we discuss the dependencies between these technologies for performing maintenance 4.0 tasks. Furthermore, we present a comparative study among the most commonly used machine learning algorithms for PdM. Our discussion is distinguished by its careful analysis and thoughtful selection of comparison criteria, which takes into account the specific needs and challenges of maintenance in industry 4.0, to ensure a comprehensive and effective comparison.

The paper is structured into seven sections as follow: in section 2 we will introduce Industry 4.0 technologies, classify them, and discuss their inter-connectivity and integration into maintenance processes. Following this, we will present an entry point to predictive maintenance based on machine learning. In Section 3 we will discover PdM required data. Section 4 discuss predictive maintenance approaches as well as we will present machine learning algorithms used in PdM problems with a summarized table that contain the advantages and limits of each of them. Section 5 presents a state-of-the-art review of current research on machine learning for predictive maintenance. The results of extracted criteria are summarized in a table. Section 6 discuss the reason for selecting such criteria for our comparative study and how can we make the best choice of an algorithm. Section 7 concludes this study work.

Proceedings of M. F. E. G. I. M. - 2022/2023

2. Maintenance 4.0

Maintenance 4.0 or the forth maintenance revolution involves the integration of Industry 4.0 technologies, to enable predictive maintenance, enhance maintenance decision-making, and optimize the maintenance process [8],[9].

2.1. Industry 4.0 Technologies

Finding a way to connect the current processes is the main problem presented by Revolution 4.0. It entails leveraging technology to connect people, operations, equipment, and machines. With the help of these 4.0 technologies, operations control and monitoring are made simpler, as is maintenance. A number of research papers [2], [8], [10], [11], [12], [13], [14], [15], [16], [17], [18] agree that various technological developments can enhance maintenance operations, whether utilized independently or in conjunction with each other. The following provides a brief description of the Industry 4.0 technologies:

- Intelligent Sensors: These sensors convert physical or chemical measurements into digital signals that provide information about factors like temperature,pressure,etc. They are regarded as the true enablers of industrial digitalization .They serve as the exclusive data sources for PdM [15].
- Big Data: According to[16]it refers to massive data that could not be scaled, extended, authorized, shared, analyzed, and calculated by standard data tools within a tolerable time frame.
- Internet of Things:It is a network that links cyber-physical systems (CPSs) and offers interconnection and interaction between physical objects, enabling the automatic gathering and retrieval of a massive flow of data [13].
- Cloud Computing: Refers to the delivery of computer services, over the internet in order to facilitate quicker innovation, adaptable resource use, and scale economies.It plays a crucial role in both on-demand computation and storing data collected by CPSs [11] .
- Cyber-physical systems: Is a system in which computer, communication, and control systems (cyberspace) are closely interwoven with both natural and artificial systems, contributing to improved efficiency [11].
- Artificial Intelligence: It refers to the simulation of human intelligence in machines that are designed to think and act like humans. It enables the development of smart predictive algorithms, and enhances maintenance operations[13].
- Machine learning: A subfield of artificial intelligence, pertains to the creation and advancement of algorithms that possess the ability to learn from data and make predictions or assessments based on that acquired knowledge [13].
- Deep learning (DL): A subfield of machine learning that is concerned with the development of algorithms inspired by the structure and function of the brain [13].
- Fuzzy logic: Deals with uncertainty and imprecision in reasoning and decision-making [19].The main application of fuzzy logic in maintenance 4.0 is condition monitoring.
- Digital twin: The digital twin is a digital representation of the real environment. It is constituted of three parts: real entities in the physical world, digital models in the virtual world, and connected data that links those two worlds (see figure 1) [16].

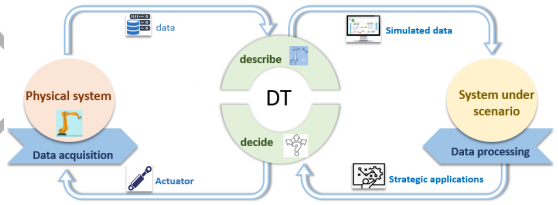


Figure 1: Digital twin process

- Simulations: In terms of technology, simulations refer to digital tools that can help predict and design production systems [11].
- Augmented reality: Overlays digital data onto the real world, enhancing human-machine interaction (see figure 2) [14]. It enables real-world augmentation and interventions by seamlessly blending virtual images with the user's perception of the environment, facilitating tasks such as repair and maintenance [13].

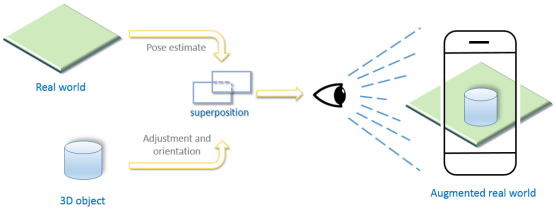


Figure 2: Augmented reality process

Proceedings of M.E.G.I.M. - 2022/2023

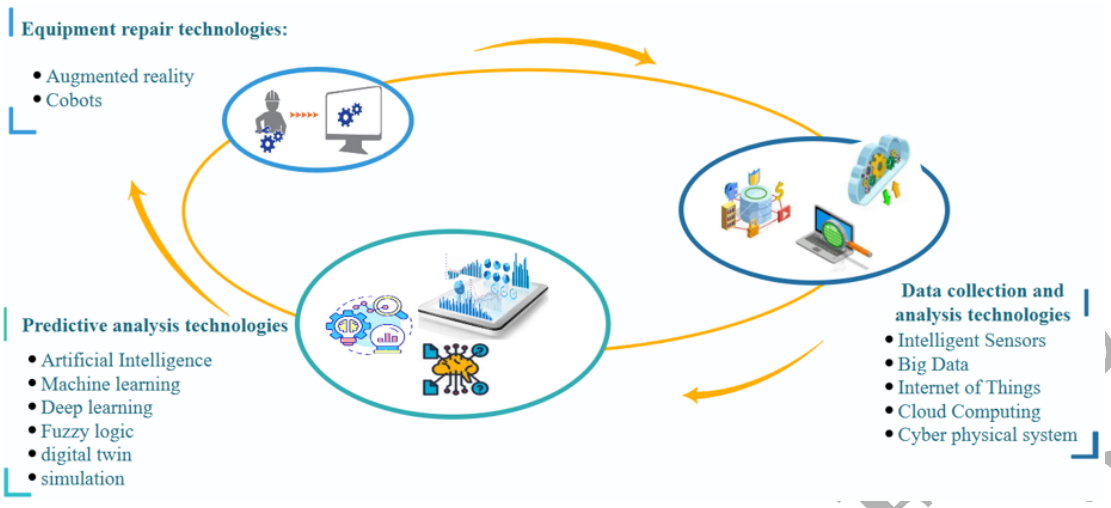


Figure 3: Classification of industry 4.0 technologies

- **Cobots: Robots for collaborative interaction are generally named "cobots." Cobots are robots that communicate with other robots and humans [11].**

These technologies can be classified according to their main functions into data collection and analysis technologies, predictive analysis technologies and equipment repair technologies. Figure 3 represents this classification:

2.2. Fusion of Industry 4.0 Technologies in the Maintenance Process

An intelligent predictive maintenance project typically requires the following tasks: creation of a failure database; data processing and cleaning; modeling, simulation, and experimental validation; instrumentation of the system to be studied; analysis of the primary failure modes; decision support; and foresight of cyber-attacks [13]. Until now, current literature research has indicated that, in order to carry out these tasks, businesses need to consider maintenance 4.0 technologies systemically. These technologies are important because they both make interconnection possible and give the new production systems' intelligence [2]. According to [11], IOT and cloud computing are currently utilized in maintenance tasks as the most appropriate technologies.

In a cyber-physical system, physical devices are linked to the internet, allowing for massive amounts of data to be collected by intelligent sensors, transmitted through the IoT, and stored in the cloud. IoT aims to address communication problems among all the systems and objects in a factory, while cloud services enable simple sharing and access to data and services from any location with an internet connection. The IoT and cloud combination allows various devices to be connected and capture enormous amounts of data [2]. As a result, the big data era has arrived [16]. With the AI, ML, and DL algorithms, we can analyze massive data and monitor and anticipate equipment failures, overloads, and other issues [2]. This makes it possible to perform predictive maintenance, which minimizes downtime resulting from unanticipated faults. Additionally, digital twins open the way for cyber-physical integration since they enable the control of the state and behaviors of physical products as well as the prediction of faults and the remaining useful life of the machine [16]. Based on simulation, the maintenance process in the virtual factory is modified to achieve optimal performance. The combination of big data and analytics can help to self-organize the maintenance process and optimize decision-making activities [2].

Before starting maintenance and repair operations, a getting started guide of maintenance strategies would be carried out in a virtual world based on virtual reality and augmented reality [16]. Then cobots could be used to support workers activities, execute maintenance tasks, and perform inspections. In maintenance, virtual reality accelerates worker training by providing an immersive simulation of maintenance operations, while augmented reality supports employees with interactive and real-time guidance for the steps required for the activities to be completed [2].

Proceedings of M. I. G. I. M. - 2022/2023

The transition to PdM 4.0 is a monumental task that will have an impact on many aspects, including security [13]. While it appears that cyber security technology is irrelevant for maintenance purposes [11]. But it is an essential aspect of modern-day business operations, including maintenance, to protect the exchange of data and critical information during maintenance operations.

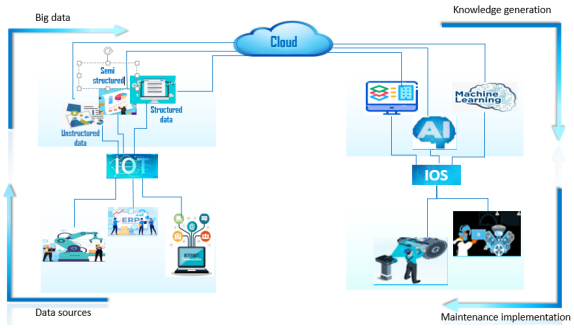


Figure 4: The fusion of industry 4.0 technologies in maintenance

2.3. Exploring Machine Learning-Based Predictive Maintenance in Industry 4.0

The predictive maintenance success prediction problem can be challenging, depending on the complexity and variability of the equipment being monitored; It is critical to take into consideration the equipment's operational importance before using PdM [7], on the other hand, the quality and availability of data, and the accuracy of the models used to predict maintenance needs, since, the lack of failure-related data is one of the main challenges data scientists have when developing these models [1]. Other factors that can impact the success of predictive maintenance include the level of expertise and resources available to implement and maintain the predictive maintenance program, the effectiveness of communication and collaboration among teams involved, and the ability to continuously adapt and improve the predictive models based on new data and insights.

Maintenance issues can be completely different in nature and the predictive information to be fed to the PdM module has in general to be tailored to the particular problem at hand, thus justifying the presence in the literature of many different approaches to PdM [20]. The choice of an appropriate approach is made according to the predefined purposes and target objects [7]. According to [21], [5] the 3 main classification approaches used for prediction are Knowledge based modeling, Physics-based modeling, and Data-driven, or Data science modeling approach. In addition the classification, we found other hybrid definitions for example cloud based [22], ML-based [2], DL-based [2]. However, PdM-related solutions based on ML techniques seem to be among the most popular [1]. Here, we highlight the ML-based approach.

3. Data Processing, Sources, and Types for Predictive Maintenance

Predictive maintenance relies heavily on the collection, storage, and analysis of big data. Big data is essential for predictive maintenance, as it provides the raw materials necessary to identify patterns and predict equipment failure. Predictive maintenance and IoT are closely related because IoT sensors can provide real-time data on equipment performance. The use of big data and IoT technologies and machine learning algorithms allows for more accurate and timely predictions .In the other hand, Big Data and Cloud computing, regardless of the model or deployment option chosen, are playing a major role in storing and analysing data. Overall, the success of predictive maintenance in Industry 4.0 relies heavily on the quality and quantity of data sources available. The convergence of those 3 fields plays the main role in PdM, with the right data sources and advanced analytic techniques, organizations can effectively predict and prevent equipment failures, resulting in improved efficiency, reduced downtime, and increased profitability [22],[3],[23].

To make maintenance decisions that are optimal, it is necessary to implement a new strategy that involves integrating various data sources required for analysis (including real-time and off-line data), and utilizing this data to construct models for predicting machine failure [12]. According to [12],[4] essential data required for the predictive maintenance are :

- Fault history: is crucial to include a sufficient number of examples of both normal operations and failures in the training data.
- Maintenance/repair history: This history in-

Proceedings of M. F. E. G. I. M. - 2022/2023

cludes information such as replaced components and preventive maintenance tasks carried out.

- Machine conditions: the data used must include time-varying functions that capture the machine's aging patterns and any anomalies that may cause a decline in performance.
- Machine data: refers to the historical operational data of assets, which may contain details on the status of the machine, important parameter names and values, uptime, downtime, and alert indicators.

These data are collected from 3 main sources (see figure 5) [16]:

1. Manufacturing resources data: including equipment data collected from smart factories by the Industrial IoT technologies, material and product data collected from themselves and service systems, and environmental data such as humidity and air quality.
2. Management data: from manufacturing information systems like enterprise resource planning (ERP) and computer aided systems (CAD),computer-aided engineering (CAE), and computer-aided manufacturing (CAM)
3. External data: including weather patterns, supplier data, etc

Data processing is a critical part of predictive maintenance. Predictive maintenance requires large amounts of data to be collected from a variety of sources, including sensors, equipment logs, maintenance records, and other sources. The data collected often contains errors and missing values. That's why data preprocessing is the next step, it involves cleaning, normalizing, and transforming the data to ensure that it is accurate and reliable [2]. Clean data is then integrated and stored for data sharing across all levels of the business. Next, cloud computing offers a platform to process and analyze real-time or offline data using sophisticated analysis methods and tools like machine learning [16], [7]. With this technology, it is possible to train machine learning algorithms on large datasets to identify patterns that may suggest potential equipment malfunction, the processes of designing a ML model proposed by [6] is presented in figure 5.

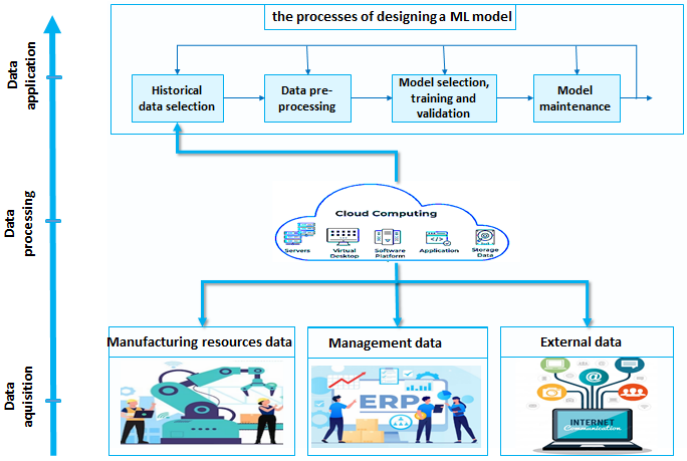


Figure 5: Data sources and the process of designing a ML model for PdM

4. Advanced Machine Learning Algorithms for Predictive Maintenance

Predictive maintenance is often subdivided into three main steps: diagnosis, prognosis, and Remaining Useful Life (RUL). Diagnosis allows to evaluate the current state of the equipment, prognosis provides for predicting future failures, and estimation of the useful life enables to plan preventive maintenance. When applying ML to perform this process we are in the presence of two types of problems and thus two types of formulations depending on the type of the output 'y'. Regression-based formulations are typically used when the output variable 'y' takes on continuous values, while classification-based PdM formulations are used when the goal is to distinguish between healthy and unhealthy system conditions, which occurs when the output variable 'y' takes on categorical values [7]. Therefore the diagnosis is considered as a classification problem, RUL is a regression problem and the prognosis can be a classification or regression problem, depending on whether the objective is to predict a class or a numerical value for an output variable.

In the following, the most frequently used ML methods to perform PdM are presented in order to make a comparative study between them:

4.1. Random Forest (RF)

RF is a supervised machine learning technique applied for classification and regression problems. It is an en-

Proceedings of M. I. G. I. M. - 2022/2023

semble learning technique that builds numerous decision trees, each of them is trained on a random subset of the data and a random subset of the features, therefore reducing overfitting and boosting the ensemble's diversity (see figure 6). The final prediction is then made, either by a majority vote for classification tasks or by averaging the forecasts of all the trees in the forest for regression tasks [24]. According to [6], when the number of variables is more than the number of samples, RF shaving performs well. However, learning can be challenging when there is a lot of data and a lot of missing data [25].

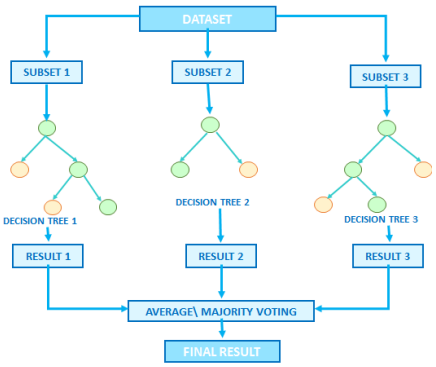


Figure 6: Structure of a RF

4.2. Support Vector Machine (SVM)

These are the newest supervised machine learning methods [26], which are employed in classification and regression problems. SVM deals with a generalized linear model using the hypothesis space of a linear function in a highdimensional feature space by creating an optimal partition hyper plane [21]. The hyperplane is selected such that it maximizes the margin between the two classes, where the margin is defined as the distance between the hyperplane and the nearest data points from each class (see figure 7). One of SVM's primary advantages is its high accuracy in classifying data and its capacity to identify the ideal point for classifying data [6].

SVMs use a method known as the kernel trick to transform the data into a higher-dimensional space where it can be more easily divided by a hyperplane in the case of nonlinear issues. In this scenario, SVM builds n-dimension hyperplanes that partition data idealistically into n classes [6]. An SVM model's complexity is unaffected by the number of features in the training data, which makes it an excellent choice for experiments when there are less instances to train than features [27], [20].

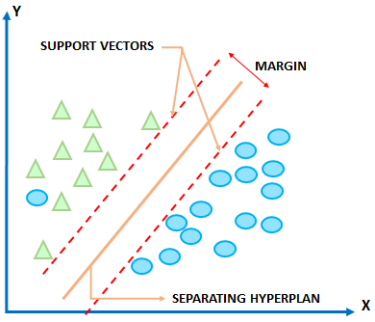


Figure 7: SVM graph

4.3. Artificial Neural Network (ANN)

ANN, or Artificial Neural Network, has gained significant importance in fault diagnosis and prognosis in the past 3 decades [2]. It is a machine learning model inspired by the human brain's structure and function, consisting of interconnected neurons that process and analyze information using activation functions (see figure 8). The network architecture is crucial for performance and depends on the problem at hand [2]. ANNs can be used for various tasks, such as classification, regression, and pattern recognition. They are trained using backpropagation to adjust neuron weights and minimize error.

On the one hand, ANNs offer several advantages such as not requiring expert knowledge for decision-making, being robust even when data is inconsistent, and enabling real-time deployment of accurate models without changing their architecture at every update. On the other hand, ANNs also have some disadvantages, including their ability to reach conclusions that contradict established rules and theories, their time-consuming training process, their "black box" nature that makes it impossible to understand their output reasoning, and their need for large datasets to learn accurately. Additionally, ANNs can be computationally expensive to train [12].

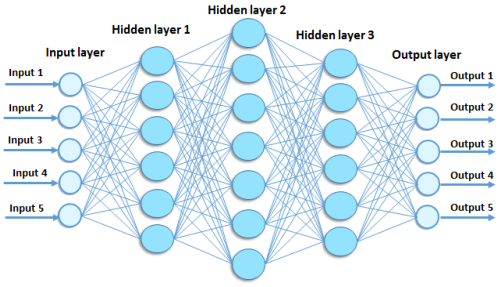


Figure 8: Structure of ANN

Proceedings of M. F. E. G. I. M. - 2022/2023

4.4. K-Nearest Neighbors (KNN)

The k-nearest neighbors algorithm is a non-parametric classification algorithm frequently employed in machine learning. It is a type of instance-based learning where predictions for a new data point are based on its similarity to the training data [2]. The algorithm selects the K closest neighbors based on a distance metric (such as Euclidean or Manhattan distance) and predicts the class label of the new data point based on the majority class among the K neighbors (see figure 9).

KNN can be used for both classification and regression tasks, but it is primarily used for classification problems. Although KNN is easy to implement and does not require a training phase, it can be sensitive to outliers and noisy data, and it can be computationally expensive for large datasets. This can be a limitation for real-time predictive systems. However, k-NNs are widely used in industrial applications, particularly in pattern recognition problems or recommendation systems [21]. They are simple, efficient, and do not require any assumptions about the data [21]. Additionally, k-NN takes more time in the testing phase, while taking less time than techniques like decision trees, neural networks, and Bayesian networks during the training phase [27].

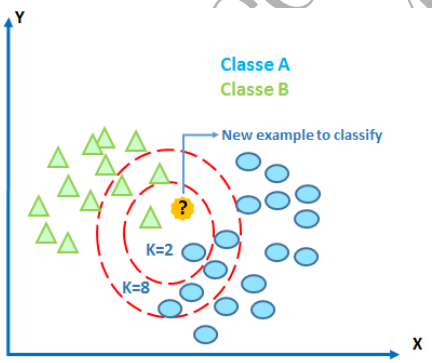


Figure 9: KNN model

4.5. Decision Tree (DT)

A Decision Tree is a non-parametric supervised method used for classification and regression [20]. DT typically consists of a number of branches, one root, a number of decision nodes, and leaf nodes (see figure 10). Each internal node of the tree represents a test on a specific feature or characteristic, and each branch corresponds to one of the possible outcomes of that test. The result of the model in light of the input data is represented by the leaf nodes [26]. The input space is recursively divided into smaller and smaller sections based on the value of a chosen feature or attribute at each node, creating a hierarchical model [27].

Decision-making situations with numerous input variables or complicated interactions between variables are particularly well-suited for decision-making by applying decision trees. DT results are simple and easy to understand (easily interpretable). However, the method might produce trees that are extremely complicated, leading to the overfitting issue. In addition, DTs performs weakly and is unstable when compared to other ML algorithms [25].

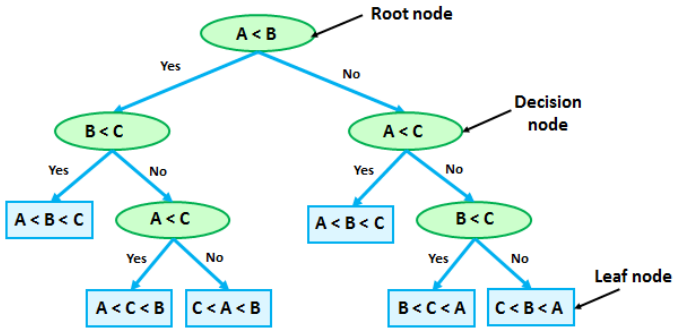


Figure 10: Structure of DT

The advantages, disadvantages, and typical applications of the widely employed conventional ML techniques for PdM are summarized in the table below.

Proceedings of M. F. E. - 1. - /2023

Algorithm	Learning strategies	Advantages	disadvantages
• RF	• Ensemble Learning • Decision Trees	• Handle high-dimensional data. • Robust to outliers and noise • Handle imbalanced data. • Less prone to overfitting compared to decision trees.	• Overfit with noisy data. • Interpretability can be an issue. • Computationally expensive, especially with large datasets
• SVM	• Binary Classification • Kernel Trick	• Effective in high-dimensional spaces. • Robust against overfitting. • handle non-linearly separable data. • Work well with even unstructured and semi structured data	• Requires careful selection of kernel and regularization parameters. • Computationally expensive. • No standard for choosing the kernel function. • Low efficiency for big data.
• ANN	• Back propagation • Gradient Descent	• Handle complex, non-linear relationships in data. • Highly adaptable and can learn from different types of data. • used for both classification and regression problems. • High classification and prediction accuracy.	• risk of overfitting (Easy to over-fit) • many weight parameters need to be trained • require greater computational resources (Computationally expensive to train) • No physical meaning • No standard to decide network structure
• KNN	• Distance Metrics	• Simple to implement and understand. • Handle non-linear relationships. • No training required. • Few parameters to tune	• Sensitivity to outliers and noisy data. • Slow and computationally expensive for large datasets. • K must be determined in advance
• DT	• Decision trees	• Easy to understand and interpret • Able to handle both categorical and numerical data • Able to handle multi-class problems • Requires little data preparation • Able to handle missing values • Fast and efficient for training and classification • Non-parametric	• unstable • easily overfitting • Biased towards variables with many levels • Does not handle well with noisy data • Poor prediction accuracy

Table 1: Learning strategies, advantages and disadvantages of RF, SVM, ANN, KNN, DT

5. State-of-the-Art in Machine Learning Applied to Predictive Maintenance

The field of machine learning for predictive maintenance has seen significant research interest in recent years, with numerous studies exploring the use of various algorithms and methodologies for fault diagnosis, tool condition assessment, and life prediction.

In papers [28],[29], and [30], a comprehensive analysis was conducted to review the most commonly used machine learning algorithms for classification and regression problems. The paper [28] thoroughly discusses the advantages and disadvantages of the presented algorithms, including a comparison of their performance and learning rates. The significance of machine learning techniques for IoT data analysis is highlighted in a paper [30]. It emphasizes the frequent use of linear regression and support vector machines for predicting and classifying high-speed sequential data.

In the cited studies [20],[26], [31], [32], [33], and [34], a comparative analysis of selected machine learning algorithms was conducted, incorporating multiple evaluation criteria. The emphasis was placed on understanding the contextual conditions under which a specific method can outperform others in a given application problem. The concept of leveraging the strengths of one method to complement the weaknesses of another was also addressed by the authors of [20], [26], and [31]. Moreover, the author of [26] recommended considering a distributed processing environment for large datasets, the results highlighted SVM as the algorithm exhibiting the highest precision and accuracy. The research further indicated that the time required to build a model and its precision (accuracy) are factors to consider, while factors like the kappa statistic and mean absolute error (MAE) play a role as well. Notably, the findings of [34] showcased SVM's superior performance compared to other classifiers, emphasizing the importance of robust and distinctive features in training classifier models. Additionally, [33] focused on the fundamental concepts, advantages, and disadvantages of each technique, concluding with practical applications to assess their performance.

Proceedings of M. F. E. G. I. M. - 2022/2023

Various evaluation measures, such as sensitivity and specificity, were employed, revealing that no single measure can provide a comprehensive assessment of classifier performance, and no single classifier can fulfill all criteria.

In the referenced papers [24], [35],[1], and [36], various methodologies are proposed for applying machine learning to predictive maintenance. The paper [24] introduces a diagnostic and prognostic model capable of predicting classes for each model, with good accuracy achieved by multiple algorithms. However, the highest accuracy, precision, and recall were obtained using random forests. [35] presents a method for tool condition assessment and life prediction based on nonlinear feature reduction and support vector regression. The paper [1] presents a novel multiple classifier PdM system (MC PDM) for integral type faults, allowing dynamic maintenance policy adjustments based on current costs and manufacturing production needs. The proposed methodology, demonstrated in a semiconductor manufacturing Ion-Implanter maintenance task, outperforms classical PvM approaches and a single SVM classifier distance-based PdM alternative. The study also highlights the superior performance of SVMs over k-NN classifiers when implementing MC PdM, with MC-PdM-knn consistently outperforming PvM approaches. [36] describes a system that combines supervised models trained on historical data with Auto-Regressive Integrated Moving Average (ARIMA) model, providing predictions for the same dataset.

In the reviewed papers [6], [25], and [7], the focus is on machine learning and reasoning for predictive maintenance, aiming to address key issues and challenges in this field. [6]presents a systematic literature review that focuses on two scientific databases, providing a comprehensive overview of machine learning techniques, their outcomes, challenges, and opportunities in the context of predictive maintenance. This review serves as a valuable foundation for new research endeavors in the PdM domain. [25] identifies major questions and technical challenges encompassing data-based modeling and fusion, such as dealing with highly imbalanced data, nonstationarity, heterogeneity of information, and knowledge transferability across tasks. Additionally, the application of these techniques to industrial setups raises concerns about cost efficiency, flexibility, and specialized training requirements. [7] highlights the results of the systematic literature review, revealing that no single algorithm can effectively address all scenarios encountered in different companies. The challenges posed by big data are also emphasized, as predictive maintenance heavily relies on large datasets, requiring further exploration of scalability, latency, and data security aspects.

In-depth analysis was conducted on pertinent research papers, extracting the essential criteria for evaluation. The findings and corresponding information have been consolidated and organized in Table 2 for convenient reference and comprehension.

6. Discussion

After analyzing the table, we found that some criteria are closely related to each other. For example, robustness is linked to an algorithm's ability to handle difficult data, which has a relationship with Tolerance to irrelevant attributes, Tolerance to redundant attributes, and Tolerance to highly interdependent attributes. These criteria all concern an algorithm's ability to handle useless or redundant attributes. On the other hand, flexibility is linked to an algorithm's ability to handle different types of data and tasks. Dealing with discrete/continuous data and Data type are two criteria that concern the type of data being manipulated. Another closely related criterion is data size and scalability, which both relate to an algorithm's ability to handle large quantities of data.

The various comparison criteria that should be chosen using scientific research are now discussed after the ML algorithms were described in the previous part, along with the advantages and disadvantages. The graph below represents the selected criteria:

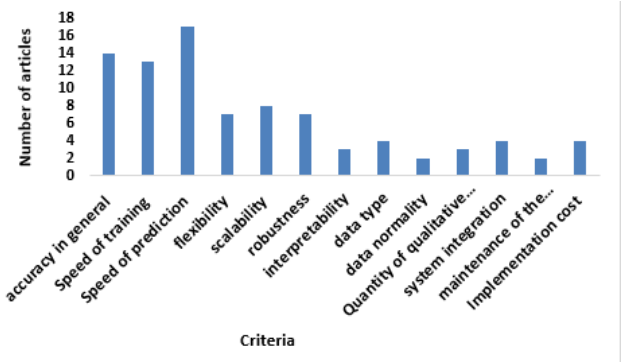


Figure 11: the selected criteria

Proceedings of M. F. E. G. I. M. - 2022/2023

reference	Accuracy in general	Speed of training	Speed of prediction	Tolerance to missing values	Tolerance to irrelevant attributes	Tolerance to redundant attributes	Tolerance to correlated features	8Dealing with discrete/continuous	Tolerance to noise	Dealing with danger of overfitting	Attempts for incremental learning	Explanation ability	Support for multi-classification	Statistics Kappa	Type of algorithm	Parametrization	Memory size	Data type	Flexibility	Data size	scalability	Robustness	Implementation
[20]	×	×	×	×	×	×	×	×	×	×	×	×				×							
[24]	×		×																				
[26]	×	×	×											×									
[28]		×	×					×		×			×		×		×	×	×	×	×		
[29]	×	×	×	×	×	×	×					×								×	×	×	×
[6]	×	×	×	×	×	×	×		×				×							×	×	×	×
[4]	×		×										×										
[36]	×		×																				
[31]	×	×	×	×	×	×	×	×	×	×	×	×	×										
[32]	×	×	×	×	×	×			×	×												×	×
[37]		×	×						×					×									
[38]	×		×																	×			
[34]	×	×	×																				
[33]	×	×	×	×	×	×	×		×	×				×	×			×		×	×	×	
[39]	×	×	×																				
[30]	×	×	×	×	×	×	×								×						×	×	×
[40]		×	×							×					×	×	×						

Table 2: A summary of the state of the art in ML applied to PdM

Table 3 shows a comparative study of commonly used ML Techniques in PdM with the 13 considered criteria.

To begin with, let us consider the criteria related to the algorithm. With the emergence of the new industrial revolution and the widespread adoption of Cyber-Physical Systems and IoT, there is a constant influx of large volumes of data. Hence, it is desirable to have an algorithm that can maintain high performance even when the input data size is increased. Such an algorithm is typically referred to as being scalable. The scalability of an algorithm is often assessed in terms of its training time and memory capacity. Therefore, algorithms that possess this characteristic are highly preferred in these dynamic and data-intensive environments.

In the context of industrial machinery, complexity and unique failure patterns are commonplace. Therefore, it is imperative that ML algorithms possess the flexibility and adaptability required to accommodate a range of machines and their specific failure patterns. This is particularly critical when comparing different algorithms. Moreover, industries face the challenge of unexpected and abnormal conditions that can impact the quality of data, for example,through sensor degradation, resulting in missing or noisy data. Thus, the robustness of algorithms is a key consideration in addressing such challenges.

When considering the output of machine learning algorithms, two crucial criteria must be taken into account: precision and interoperability of results. Precision is generally evaluated using metrics such as Mean Absolute Error, Mean Squared Error (MSE), and Mean Relative Error (MRE). On the other hand, the ability to interpret the output of an algorithm is equally important. An algorithm that generates easily understandable results enables maintenance teams to comprehend the reasons behind a given prediction and develop maintenance strategies accordingly. However, choosing the right algorithm for the amount of data available can significantly impact the accuracy and effectiveness of the model, as well as the success of the predictive maintenance project as a whole. For instance, if the quantity of data is lower than 100,000, any machine learning model can be used [28]. But if the quantity of data exceeds that threshold, the descent of gradient is preferred. Hence, to ensure long-term precision and relevance, machine learning algorithms must be designed with ease of maintenance and update in mind, taking into account both the amount

Proceedings of M. F. E. G. I. M. - 2022/2023

criteria	RF	SVM	ANN	KNN	DT
Accuracy in general	Excellent	Excellent	Excellent	V.good	Good
Training speed	V.good	Average	Average	V.good	V.good
Prediction Speed	Excellent	Excellent	Excellent	Good	V.good
Flexibility	V.good	V.good	High	Excellent	Excellent
Scalability	Excellent	Good	Excellent	Good	Average
Robustness	Excellent	Excellent	Good	Good	Average
Interpretability of results	V.good	V.good	Good	V.good	Excellent
Data type	Both	Structured	Both	Both	Structured
Data normality	Does not require normally distributed data	Does not require normally distributed data	Require normally distributed data for optimal performance	Does not require normally distributed data	Does not require normally distributed data
Quantity of qualitative and quantitative variables	Can handle both qualitative and quantitative variables	Best performance with quantitative variables	Can handle both qualitative and quantitative variables	Can handle both qualitative and quantitative variables	Best performance with quantitative variables
System integration	Easy to integrate with other systems	Require additional software	Require additional software	Easy to integrate with other systems	Easy to integrate with other systems
Maintenance of the algorithm	Average	Average	Good	Average	Average
Implementation cost	Moderate	High	High	Low	Low

Table 3: the evaluation of the algorithms according to the selected criteria

of available data and the Interpretability of the results.

Another key criterion to consider is the speed of a machine learning algorithm to perform its tasks, which is dependent on two other factors: training time and prediction time. This criterion is of paramount importance since maintenance teams require adequate time to plan and execute preventive maintenance tasks. Thus, the ability of an algorithm to operate efficiently and rapidly is crucial in ensuring the timely execution of maintenance activities.

Several criteria must be considered when evaluating the characteristics of data. Firstly, the type of data must be assessed, as classical ML algorithms are typically applicable only to structured data. In contrast, for unstructured data, NNs are often preferred. Secondly, the quantity of qualitative and quantitative variables in the dataset is critical. For instance, certain models such as DTs are not effective when quantitative variables are dominant or linear relations exist between these variables, resulting in a high risk of overfitting when additional data is added. Additionally, the normality of the data is another important criterion to consider. It is worth noting that parametric models such as NNs and Naïve Bayes (NB) perform well on data that follow normal probability laws [30].

Two key criteria that must be taken into account when plementing a machine learning model are system integration and implementation cost [41]. It is imperative that ML algorithms can interface with real-time data acquisition and control systems to collect relevant input data and provide actionable outputs for maintenance teams. By doing so, the costs and development times associated with ML algorithm implementation can be reduced. Moreover, it is necessary to consider the cost of software, necessary infrastructure, and employee training during the implementation process to ensure the overall success of the model.

Ultimately, the choice of which machine learning model to use depends on the specific requirements and characteristics of each company [33]. After determining the project requirements and using the previously presented table, the weighted voting method can be applied to select the most suitable model for the situation at hand. For example, if a company places a higher value on the interpretability of results, they can assign a higher weight to this criterion in the decision-making process. Similarly, if a company has a large amount of unstructured data, they can assign a higher weight to the suitability criterion for unstructured data.

After assigning weights to the different evaluation criteria, each machine learning model can be evaluated based on each criterion, and a score can be calculated for each model. The weighted scores for each model can then be added together to determine the overall score for each model. Finally, the model with the highest overall score can be selected as the most suitable

Proceedings of M. I. E. G. I. M. - 2022/2023

model for the specific requirements and characteristics of the company.

7. Conclusions

The emergence of the new industrial revolution has led to the development of various technologies that can effectively address the challenges of predictive maintenance. Also, the fact that maintenance represents 15 to 60% of the total operating cost of all manufacturing emphasizes how to benefit from these technologies and use them in the field of maintenance. They changed the way of doing things and took maintenance to another level, and the term maintenance 4.0 appeared.

It is important to keep in mind Wolpert and Macready's well-known No Free Lunch Theorem (NFL) [42] when considering the selection and evaluation of machine learning algorithms in the field of PdM. This theorem states that no algorithm is universally better than another across all possible problems. Therefore, when choosing an algorithm for a specific maintenance scenario, it is essential to consider critical criteria for a model's success, such as data type, amount of data, model complexity, interpretability, system integration, and implementation cost. Our study contributes to the development of effective PdM strategies by providing valuable insights into the most suitable ML algorithms for different maintenance scenarios, and by using the weighted voting method to select the best model for a specific situation. While accuracy is an important factor in evaluating algorithm performance, it is not the sole criterion, as there are multiple measures that impact our decision, and none of the algorithms can satisfy all criteria. By keeping this in mind, companies can enhance the efficiency and effectiveness of their maintenance operations.

References

[1] Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Seán McLoone, and Alessandro Beghi. Machine learning for predictive maintenance: A multiple classifier approach. *IEEE transactions on industrial informatics*, 11(3):812–820, 2014.

[2] Yongyi Ran, Xin Zhou, Pengfeng Lin, Yonggang Wen, and Ruilong Deng. A survey of predictive maintenance: Systems, purposes and approaches. *arXiv preprint arXiv:1912.07383*, 2019.

[3] Christian Krupitzer, Tim Wagenhals, Marwin Züfle, Veronika Lesch, Dominik Schäfer, Amin Mozaffarin, Janick Edinger, Christian Becker, and Samuel Kounev. A survey on predictive maintenance for industry 4.0. *arXiv preprint arXiv:2002.08224*, 2020.

[4] Marina Paolanti, Luca Romeo, Andrea Felicetti, Adriano Mancini, Emanuele Frontoni, and Jelena Loncarski. Machine learning approach for predictive maintenance in industry 4.0. In *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, pages 1–6. IEEE, 2018.

[5] Tiago Zonta, Cristiano André Da Costa, Rodrigo da Rosa Righi, Miromar Jose de Lima, Eduardo Silveira da Trindade, and Guann Pyng Li. Predictive maintenance in the industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, 150:106889, 2020.

[6] Thyago P Carvalho, Fabrízzio AAMN Soares, Roberto Vita, Roberto da P Francisco, João P Basto, and Symone GS Alcalá. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137:106024, 2019.

[7] Jovani Dalzochio, Rafael Kunst, Edison Pignaton, Alecio Binotto, Srijnan Sanyal, Jose Favilla, and Jorge Barbosa. Machine learning and reasoning for predictive maintenance in industry 4.0: Current status and challenges. *Computers in Industry*, 123:103298, 2020.

[8] Mounia Achouch, Mariya Dimitrova, Khaled Ziane, Sasan Sattarpanah Karganroudi, Rizck Dhouib, Hussein Ibrahim, and Mehdi Adda. On predictive maintenance in industry 4.0: Overview, models, and challenges. *Applied Sciences*, 12(16):8081, 2022.

[9] Małgorzata Jasiulewicz-Kaczmarek and Arkadiusz Gola. Maintenance 4.0 technologies for sustainable manufacturing -an overview. *IFAC-PapersOnLine*, 52-10:91–96, 11 2019.

[10] Alejandro Germán Frank, Lucas Santos Dalenogare, and Néstor Fabián Ayala. Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210:15–26, 2019.

Proceedings of M. E. E. G. I. M. - 2022/2023

[11] Abhijeet Ghadge, Merve Er Kara, Hamid Moradlou, and Mohit Goswami. The impact of industry 4.0 implementation on supply chains. *Journal of Manufacturing Technology Management*, 31(4):669–686, 2020.

[12] Go Muan Sang, Lai Xu, Paul de Vrieze, Yuewei Bai, and Fangyu Pan. Predictive maintenance in industry 4.0. In *Proceedings of the 10th International Conference on Information Systems and Technologies*, pages 1–11, 2020.

[13] Emiliano Traini, Giulia Bruno, Gianluca D'antonio, and Franco Lombardi. Machine learning framework for predictive maintenance in milling. *IFAC-PapersOnLine*, 52(13):177–182, 2019.

[14] Luca Silvestri, Antonio Forcina, Vito Introna, Annalisa Santolamazza, and Vittorio Cesarotti. Maintenance transformation through industry 4.0 technologies: A systematic literature review. *Computers in Industry*, 123:103335, 2020.

[15] Luca Silvestri, Antonio Forcina, Vito Introna, Annalisa Santolamazza, and Vittorio Cesarotti. Maintenance transformation through industry 4.0 technologies: A systematic literature review. *Computers in Industry*, 123:103335, 2020.

[16] Qinglin Qi and Fei Tao. Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *Ieee Access*, 6:3585–3593, 2018.

[17] Aymane Sahli, Richard Evans, and Arthi Manohar. Predictive maintenance in industry 4.0: Current themes. *Procedia CIRP*, 104:1948–1953, 2021.

[18] Jennifer S Raj. A comprehensive survey on the computational intelligence techniques and its applications. *Journal of ISMAC*, 1(03):147–159, 2019.

[19] Antonio Forcina, Vito Introna, and Alessandro Silvestri. Enabling technology for maintenance in a smart factory: A literature review. *Procedia Computer Science*, 180:430–435, 2021.

[20] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[21] Myriam Lopez. Prédiction d'erreurs critiques pour l'aide à la maintenance industrielle. In *Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA'21) Plate-Forme Intelligence Artificielle (PFIA'21)*, pages 82–83, 2021.

[22] Martin Pech, Jaroslav Vrchota, and Jiří Bednář. Predictive maintenance and intelligent sensors in smart factory. *Sensors*, 21(4):1470, 2021.

[23] Jay Lee, Hossein Davari Ardakani, Shanhu Yang, and Behrad Bagheri. Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. *Procedia cirp*, 38:3–7, 2015.

[24] Dwi Kusumaningrum, Nani Kurniati, and Budi Santosa. Machine learning for predictive maintenance. In *Proceedings of the International Conference on Industrial Engineering and Operations Management. IEOM Society International*, pages 2348–2356, 2021.

[25] Alberto Diez-Olivan, Javier Del Ser, Diego Galar, and Basilio Sierra. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards industry 4.0. *Information Fusion*, 50:92–111, 2019.

[26] FY Osisanwo, JET Akinsola, O Awodele, JO Hinmikaiye, O Olakanmi, J Akinjobi, et al. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138, 2017.

[27] Baljinder Ghotra, Shane McIntosh, and Ahmed E Hassan. Revisiting the impact of classification techniques on the performance of defect prediction models. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, volume 1, pages 789–800. IEEE, 2015.

[28] Susmita Ray. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE, 2019.

[29] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386, 2020.

Proceedings of M. I. E. G. I. M. 2022/2023

[30] Mohammed H Alsharif, Anabi Hilary Kelechi, Khalid Yahya, and Shehzad Ashraf Chaudhry. Machine learning algorithms for smart data analysis in internet of things environment: taxonomies and research trends. *Symmetry*, 12(1):88, 2020.

[31] Hetal Bhavsar and Amit Ganatra. A comparative study of training algorithms for supervised machine learning. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(4):2231–2307, 2012.

[32] Sinem Bozkurt, Gulin Elibol, Serkan Gunal, and Ugur Yayan. A comparative study on machine learning algorithms for indoor positioning. In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pages 1–8. IEEE, 2015.

[33] Amr E Mohamed. Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied*, 7(2):1–15, 2017.

[34] Kapil Sethi, Ankit Gupta, Gaurav Gupta, and Varun Jaiswal. Comparative analysis of machine learning algorithms on different datasets. In *Circulation in Computer Science International Conference on Innovations in Computing (ICIC 2017)*, volume 87, 2019.

[35] Tarak Benkedjouh, Kamal Medjaher, Noureddine Zerhouni, and Said Rechak. Health assessment and life prediction of cutting tools based on support vector regression. *Journal of intelligent manufacturing*, 26:213–223, 2015.

[36] Ameeth Kanawaday 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 (ICSESS)*, pages 87–90. IEEE, 2017.

[37] Pankaj Ajit. Prediction of employee turnover in organizations using machine learning algorithms. *algorithms*, 4(5):C5, 2016.

[38] BK Bhavitha, Anisha P Rodrigues, and Niranjan N Chiplunkar. Comparative study of machine learning techniques in sentimental analysis. In *2017 International conference on inventive communication and computational technologies (ICICCT)*, pages 216–221. IEEE, 2017.

[39] Kajaree Das and Rabi Narayan Behera. A survey on machine learning: concept, algorithms and applications. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2):1301–1309, 2017.

[40] Vansh Jatana. Machine learning algorithms, 06 2019.

[41] Carlos Resende, Duarte Folgado, João Oliveira, Bernardo Franco, Waldir Moreira, Antonio Oliveira-Jr, Armando Cavaleiro, and Ricardo Carvalho. Tip4. 0: industrial internet of things platform for predictive maintenance. *Sensors*, 21(14):4676, 2021.

[42] David H Wolpert, William G Macready, et al. No free lunch theorems for search. Technical report, Citeseer, 1995.

Proceedings of M. F. E. G. I. M. - 2022/2023