Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DATA-DRIVEN PREDICTIVE ANALYSIS FOR CUTTING MACHINE FAILURES: A TECHNICAL REPORT ON RELIABILITY OPTIMIZATION

Salma MAATAOUI [1], Ghita BENCHEIKH [2], Maria EZZIANI[3], Ghizlane BENCHEIKH[4]

[1] PhD Student, Moulay Ismail University of Meknes, Morocco

[2] Professor, Laboratory of Lineact, Cesi Engeneering School, France

[3] Engineer Student, School of Science and Engineering, Al Akhawayn University, Ifrane, Morocco

[4] Professor, Moulay Ismail University of Meknes, Morocco

E-mail : [1]salma.maataoui@edu.umi.ac.ma, [2]gbencheikh@cesi.fr, [3]M.Ezziani@aui.ma, [4]g.bencheikh@umi.ac.ma

ABSTRACT

The prevention of recurring failures in modern manufacturing systems is of paramount importance for minimizing costs and downtime. Despite the potential for real-time data analysis using sensors offered by Industry 4.0 technologies, their widespread adoption, particularly among smaller manufacturing systems, remains a challenge. In response, this paper introduces an alternative approach to predictive maintenance planning, utilizing historical maintenance intervention data in the absence of sensor-based information. The study investigates the pivotal role of Artificial Intelligence (AI), specifically Machine Learning (ML) and Prognostics and Health Management (PHM), in augmenting the efficiency of predictive maintenance. Utilizing a comprehensive dataset from Schleuniger cutting machines spanning June 2021 to June 2023, our research evaluates two predictive maintenance approaches: Precision-Based Maintenance Prediction (PBMP) and Occurrence-Driven Maintenance Prediction (ODMP). The objective is to extract valuable insights from historical maintenance data, enabling proactive decision-making and preventing future failures. The deployment results presented in this study demonstrate the effectiveness of predicting the number of failures, providing valuable information that can enhance maintenance planning and reduce total downtime. By addressing the practical challenges faced by smaller companies in adopting sensor technologies, this research contributes valuable insights to the broader landscape of predictive maintenance in manufacturing.

Keywords: *Industry 4.0, Predictive maintenance, Artificial Intelligence, Prognostics and Health Management, Schleuniger Cutting Machines, Data-Driven Decision-Making*

1. INTRODUCTION

Maintenance planning has traditionally relied on Key Performance Indicators (KPIs) to evaluate equipment health and fine-tune maintenance strategies. Well-established KPIs like Mean Time Before Failure (MTBF), Mean Time to Repair (MTTR), Overall Equipment Effectiveness (OEE), and Asset Utilization (AU) provide valuable insights into equipment reliability, maintenance efficiency, and overall production performance. These metrics enable organizations to make data-driven decisions and effectively reduce downtime. In the modern era of Industry 4.0, the integration of advanced technologies such as Artificial Intelligence (AI), Intelligent Automation (IA), Machine Learning (ML), and Predictive Maintenance (PdM) has become increasingly essential across various industrial sectors [6, 26]. ML excels in uncovering hidden insights within data while enabling machines to learn from examples and past experiences. PdM significantly improves machine uptime by proactively preventing failures through advanced analytics applied to data collected from condition monitoring and sensors [35, 7, 13]. The PdM strategy relies on Prognostic and Health Management (PHM) techniques for predicting the Remaining Useful Life (RUL) of equipment, which has gained prominence due to its ability to early identify potential breakdowns [36, 4,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

22], thereby reducing overall maintenance costs and mitigating industrial accidents [16, 5].

Numerous research papers have explored ML's use in monitoring and predicting machine behavior, with a focus on sensor-based datasets [19, 10, 17]. However, this approach may pose challenges for smaller companies with resource constraints. Despite the transformative impact of Industry 4.0 technologies on PdM practices, their implementation is hindered by preparatory work, such as sensor installation and effective data management. Many manufacturing systems, hesitant to fully adopt these technologies, persist with traditional maintenance strategies, relying on KPIs like MTBF and MTTR. This raises questions about the adequacy of solely relying on these KPIs when sophisticated predictive maintenance models or sufficient datasets are lacking. Hence, an exploration into the standalone efficacy of MTBF and MTTR as predictive metrics for breakdowns is warranted, alongside an inquiry into the nature of information extractable from these indicators and their potential for future projections.

In response to the cited research questions, our investigation focuses on predicting machine failures by applying ML algorithms using KPIs historical data. This generic algorithm-driven approach adapts to historical data patterns, constructing logic for predictions based on previously encountered data. A notable example is found in [22], where a probabilistic risk assessment approach utilizing historical maintenance data accurately predicted post-maintenance failures during training. Leveraging the Weibull distribution, often referred to as the "bathtub curve," the study extended its analysis to testing data, facilitating precise health assessments. Integral to the study was the examination of part replacements and MTTR, instrumental in identifying preventative maintenance and gauging asset health risks. Moreover, the study established a correlation between corrective maintenance and post maintenance failure probabilities by scrutinizing failures following these actions in the testing data. In a related study, the work by [29] emphasizes the importance of factors such as reliability, availability, Time Between Failures (TBF), and Time To Repair (TTR) in ensuring the uninterrupted functionality of an industrial system. This research conducted a preliminary examination with the goal of enhancing the reliability of specific workstations within the milk powder and curd production lines at a dairy plant, with the ultimate aim of improving availability under authentic working conditions. Through the utilization of field failure and repair data, coupled with Weibull based software and fitness indexes, the study input TBF and TTR data into the software. The findings of this examination suggest that optimizing the maintenance schedule based on reliability analysis, particularly for subsystems experiencing higher failure rates, proves advantageous for both performance optimization and the preservation of dairy product quality. These two studies are prove that KPIs can be bring insightful information and improve maintenance planning.

Building upon these insights, [18] employs ML techniques to predict machine performance using KPIs, demonstrating the advantages of this approach in reducing manual efforts, analyzing machine wise KPIs (such as OEE, MTBF, and MTTR), and identifying critical machines for proactive interventions. Additionally, [2] proposes an innovative framework for real-time monitoring and prediction of OEE and related metrics using deep learning, exhibiting high predictive accuracy in an advanced automotive industry context. The study [20] utilizes AHC-CA clusters obtained through correspondence analysis, subjected to ML regression algorithms. Maintenance performance indicators, including MTBF, MTTR, and Wasted Oil per Machine-Month (WOMM), serve as independent variables for algorithm selection. The study employs non-parametric ML algorithms such as: Support Vector Regression (SVR), k-Nearest Neighbour (kNN), Decision Tree (DT), Random Forest (RF) and others.

1.1 Research Questions

This paper addresses complexities in Enterprise Asset Management (EAM), emphasizing historical data on maintenance failures with a focus on MTBF and MTTR. The research seeks answers to crucial questions:

1. Is the exclusive use of MTBF and MTTR sufficient for effective predictive analysis?

2. What insights can be derived from historical maintenance intervention data?

3. How can these insights inform predictive maintenance strategies without comprehensive labeled data?

The paper is organized into distinct sections: a historical perspective, the paper traces the evolution of production systems and the concurrent development of maintenance strategies. This historical exploration lays the groundwork for the subsequent discussion on the integration of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

advanced technologies into contemporary manufacturing and understands what are the basic KPIs, such as MTBF and MTTR. The subsequent section pivot towards the application of ML techniques, with a dedicated focus on regression and classification algorithms. In-depth elucidations of the deployed algorithms and pertinent evaluation metrics are presented to foster a nuanced comprehension. The crux of the proposed methodology encompasses two predictive maintenance approaches: Precision-Based Maintenance Prediction (PBMP), targeting Time To Repair (TTR), and Occurrence-Driven Maintenance Prediction (ODMP), designed for forecasting the number of failures. Rigorous testing and evaluation procedures are applied to various algorithms within each category. The ensuing section articulates the outcomes of these evaluations, underscoring the superior performance of the ODMP approach in predicting the frequency of failure occurrences.

1.2 Contributions

Inspired by these endeavors, our study introduces two distinct approaches—Precision-Based Maintenance Prediction and Occurrence-Driven Maintenance Prediction. These methodologies seek to comprehend limitations and extract pivotal factors from fundamental KPIs. Leveraging an extensive dataset from diverse Schleuniger Cutting machines, our investigation assesses the efficacy of two predictive maintenance approaches, aiming to identify patterns and underscore the significance of PHM in augmenting machine reliability. Specifically, we will conduct a regression analysis to predict the point at which the machine is likely to fail, and a classification study to anticipate the number of potential failures. By seeking to identify meaningful patterns, our study highlights the significance of PHM in enhancing machine reliability.

Aligned with principles of data-driven decision-making, our proactive approach aspires to enhance production efficiency and curtail maintenance costs, providing invaluable insights for optimizing Enterprise Asset Management (EAM).

Structured into five sections, the paper first details the machines analyzed, encompassing datasets and foundational components. In Section 3, we propose two distinct approaches for predicting machine failures using historical data. The subsequent section discusses the results obtained from these proposed algorithms, and Section 5 encapsulates the primary research findings and limitations. Additionally, it provides a glimpse into potential future directions and applications within the field of Production Engineering.

2. MAINTENANCE STRATEGIES IN THE EVOLUTION OF INDUSTRY: FROM RUN-TO-FAILURE TO PREDICTIVE MAINTENANCE

The industrial landscape has undergone significant transformations throughout history, driven by a series of revolutionary phases known as industrial revolutions. Each revolution has introduced groundbreaking advancements, reshaped the manufacturing sector and altering the way society produces goods [33] (Figure 1).

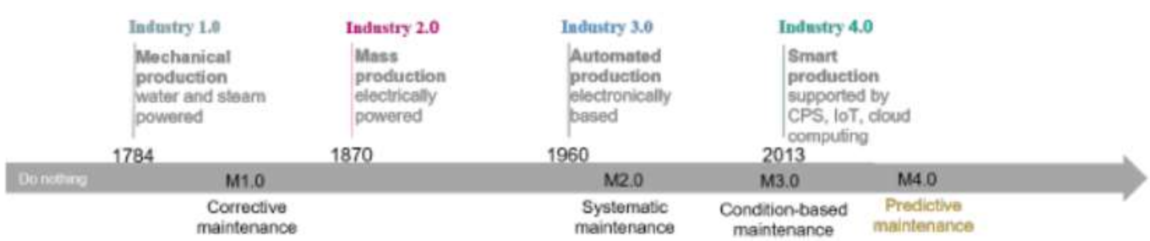

Figure 1: Production vs Maintenance evolution over time

Industry 1.0: started around 1784, ushered in a new era with the introduction of mechanical manufacturing systems powered by water and steam. This shift from manual labor to mechanized processes laid the foundation for industrialization.

Industry 2.0: emerged in the late 19th century, characterized by the widespread adoption of electrical energy. This innovation enabled mass production on a scale previously unimaginable, transforming industries and driving economic growth.

Industry 3.0: marked the onset of the modern era in the middle of the 20th century. The use of computers and technology in industries brought automation, boosting efficiency and productivity. Information and Communication Technologies (ICT) saw remarkable growth during this period, further propelling advancements.

Industry 4.0: triggered, at the beginning of the 21st century, a paradigm shifts with the integration of cyber-physical systems (CPS), the Internet of Things (IoT), and cloud computing. This revolution opened doors to smart manufacturing, where interconnected systems communicate and learn from data, leading to highly flexible and optimized production processes.

During the first and second industrial revolutions, machines were designed simply, making them easy to repair. Maintenance actions were performed after breakdowns occurred, a reactive approach known as "Maintenance 1.0", "run-to-failure", or "corrective maintenance". However, this approach resulted in significant financial losses due to the lack of anticipation, including production loss and repair costs.

As machines became more complex and breakdowns more frequent, the concept of preventive maintenance emerged. Initially, systematic preventive maintenance (Maintenance 2.0) was applied, based on MTBF and MTTF. MTBF, in one hand, refers to the average time between the occurrence of one failure and the next failure in a system or piece of equipment. Mathematically, MTBF can be calculated by dividing the total operational time of a system by the number of failures that occur during that time period (eq. 1). A high MTBF indicates that the system is less prone to failures and can operate for extended periods without encountering issues.

$$MTBF = \frac{Total\ Operational\ Time}{Number\ of\ Failures} \quad (1)$$

MTTR, in the other hand, represents the average time it takes to repair or restore a failed system or equipment to its normal operating condition after a failure has occurred. MTTR is a measure of maintainability, and a lower MTTR is generally desirable, as it indicates that the system can be quickly repaired and returned to service. MTTR is calculated by dividing the total downtime due to failures by the number of failures (eq. 2).

$$MTTR = \frac{Total\ Downtime\ due\ to\ Failures}{Number\ of\ Failures} \quad (2)$$

Both MTBF and MTTR play crucial roles in assessing the overall reliability and maintainability of systems, and they are used in various industries to support maintenance strategies and optimize operational efficiency. However, this fixed and predetermined approach does not consider the actual usage and maintenance practices during the equipment's life cycle. To improve this strategy, a new approach called Condition-Based Maintenance (CBM) or Maintenance 3.0 was proposed. CBM involved continuously monitoring the system's health and making maintenance decisions based on real-time condition data. This approach gained popularity in military and industrial sectors due to reduced maintenance and logistics costs, increased equipment availability, and protection against failures of critical equipment.

More sophisticated maintenance strategies have since been proposed, incorporating advanced methods and tools. PHM have garnered significant attention, leading to the emergence of predictive maintenance (Maintenance 4.0). PdM uses historical operating data and current condition information to predict future equipment states and anticipate breakdowns. PdM become productive using IoT-data with zero-downtime for maintenance in industries by estimating the RUL [3]. By planning only necessary maintenance tasks, this approach minimizes unexpected breakdowns and reduces the risk of accidents in workshops. The success of these advanced maintenance strategies relies on data availability, as data-driven decision-making is crucial in generating valuable insights for efficient maintenance planning and execution.

3. MACHINE LEARNING ALGORITHMS AND EVALUATION METRICS

Machine learning techniques are broadly classified into two main categories: supervised learning and unsupervised learning. The primary distinction between these categories lies in the nature of the training data. In supervised learning, models are trained using labeled data, where each observation is paired with its corresponding expected output. This approach is particularly relevant to our study, as we seek to predict an output value. In contrast, unsupervised learning involves training models with unlabeled data, aiming to discover inherent patterns or structures within the data itself, which is not pertinent to our research focus.

In the context of supervised learning, models undergo training with labeled data, denoted as "experiences" where each input x is associated with its corresponding expected output information y. These experiences are represented as (x_i, y_j), $i = 1, \ldots, n$, where $x_i \in R_p$ are feature vectors with p input parameter, n number of experiences and $y_i \in E$. These models act as opaque entities, adjusting their parameters based on these experiences. The primary objective is to discover a function f that establishes the relationship between the input x and the output y, represented as $f(x) = y$. By utilizing this function (f), we can generate predictions for new observations (x') where the output is unknown.

Supervised learning technique can be categorized into two main groups: regression (continuous modeling) and classification (discrete modeling). Both approaches are used for predictive modeling but differ in terms of the target variables (nature of the set E). In regression, the target variables are continuous, while in classification, the target variable vary in a countable range that can be Boolean set, N or a set of finite elements $\{e_1, e_2, \ldots, e_k\}$.

In the realm of predictive maintenance, various ML algorithms are deployed for forecasting maintenance needs. In the next subsections we present an overview of existing classification and regression algorithms and the associated evaluation metrics.

3.1 Classification algorithms

This section presents an overview of prominent classification algorithms utilized for maintenance prediction:

Decision Trees (DT) is a hierarchical structure that is employed for the representation of discrete functions. The process involves partitioning the dataset into smaller subsets based on features [21]. The structure can be visualized as a tree or matrix, where higher levels define sets of conditions, and lower levels specify sets of actions to be executed when certain conditions are met.

Random Forest (**RF**: is a two-step process. The first step consists of building individual trees using a bootstrap sample taken from the training dataset, involving random selection with replacement [23]. Second step consists of randomly picking the best split when constructing the tree, considering either all input features or a subset of them [14].

K-Nearest Neighbors (KNN): predicts the class or value of a data point based on the classes or values of its nearest neighbors. The "k" in KNN refers to the number of neighbors considered for the prediction [31].

Light Gradient Boosting Machine (LGBM) is rooted in decision-tree techniques. It adopts a histogram-based algorithm and grows trees leaf-wise, resulting in quicker training and reduced memory [11].

Multi-layer Perceptrons (MLPs): consist of interconnected nodes organized into layers: an input layer, one or more hidden layers, and an output layer. Each connection between nodes carries a weighted value, and neurons in the hidden layers apply transformations to the data [24]. Nonlinear activation functions, such as the sigmoid, hyperbolic tangent (tanh), or rectified linear unit (ReLU), are typically used in the hidden layers. These functions introduce non-linearity, enabling the network to learn complex relationships within the data. MLPs learn through a process called backpropagation, where the algorithm adjusts the weights of connections iteratively based on the difference between predicted and actual outputs. This process aims to minimize a chosen loss or cost function.

In evaluating classification models, various metrics are employed; each providing insights into different aspects of model performance. True Positives (TP) occur when the model correctly anticipates the positive class, True Negatives (TN) when it accurately recognizes the absence of the condition, False Positives (FP) when it erroneously suggests its presence, and False Negatives (FN) when it inaccurately indicates its absence. Precision (P), Recall (R), and F1 score are common metrics for assessing classification algorithms, with their formulations presented in Table 1.

Table 1: Evaluation Metrics for classification Algorithms

Metric	Formula
Precision (P)	$P = \frac{TP}{TP+FP}$
Recall (R)	$R = \frac{TP}{FP+FN}$
F1 score	$F1 = 2\frac{P \times R}{P+R}$

3.2 Regression algorithms

This section provides an overview of prominent regression algorithms employed for maintenance prediction.

Linear regression is a statistical method employed to establish the relationship between independent and dependent variables. It graphically represents this relationship through a linear function, known as the regression function [25]. This process involves finding a linear function f that captures the relationship between an input xi and an output y_i expressed as $y_i = f(x_i) = \beta_0 + \Sigma_j x_{i,j} \beta_j$, where i corresponds to an experience and j the index of parameters. To define this function f, it is necessary to estimate the regression parameters β_0 and β that minimize the selected evaluation metrics.

Support Vector Model (SVM) is an extension of the Support Vector Model (SVM) for regression problem solving; SVM efficiently handles tasks with high-dimensional features through nonlinear mapping and optimal decision function modeling [9]. The process involves transforming input vectors into higher-dimensional spaces via predetermined nonlinear mapping and using limited sample data to train an optimal decision function model.

Gradient Boosting Regression (GBR) is an iterative technique solving both classification and regression problems by creating and refining weak learners, such as decision trees, at each stage [32].

Lasso Regression (LR) stands as a robust method for examining high-dimensional data by enhancing predictive accuracy, facilitating feature selection and ensuring interpretability within a computationally efficient framework [8].

General Regression Neural Network (GRNN) approach leverages neural networks to establish intricate relationships between input features and output variables. The neural network structure comprises interconnected nodes, resembling the neurons in the human brain, organized into layers—input, hidden, and output layers [28].

Random Forest Regression (RFR) constructs a set of decision trees during the training phase. Each decision tree independently learns patterns and relationships within the data. It introduces randomness in the model-building process by considering only a subset of features and data samples for each tree. The predictions from trees are aggregated to produce the final prediction of the Random Forest. This ensemble approach mitigates the risk of overfitting by reducing sensitivity to noise and outliers present in the dataset [27].

Ridge Regression, also known as Tikhonov regularization or L2 regularization, is a linear regression technique used for predictive modeling in situations where the presence of multi collinearity among predictor variables which can lead to instability in traditional linear regression models [15]. This method have has outperformed existing methods in prediction accuracy and embedding quality through experiments published in the paper [30].

To assess the performance of regression algorithms in predictive maintenance, key evaluation met rics include Mean Square Error (MSE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2), as summarized in Table 2.

Table 2: Evaluation Metrics for Regression Algorithms

Metric	Formula
Mean Square Error (MSE)	$MSE = \frac{1}{n}\sum_i (y_i - \hat{y})^2$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{MSE}$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_i \lvert y_i - \hat{y} \rvert$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2}$

Where n is the sample size, y_i are the actual values to be predicted, $\bar{y}$ the mean value of y_i and $\hat{y}$ are the estimated values.

4. METHODOLOGY AND APPLICATION

In this section, we elucidate the systematic methodology employed in our study to glean insights into predictive maintenance for Schleuniger Cutting machines. This methodology encompasses data collection, feature selection, and the application of distinct predictive maintenance approaches for both precision-based and occurrence-driven.

4.1 System description

The studied system is a cutting machine: Crimp Center 64 SP from Schleuniger group (figure 2). These wire cutting machines exhibit a high level of proficiency in performing precision cutting, crimping, and sealing tasks, ensuring the consistent delivery of top-quality results [12].

Figure 2: Cutting machine Crimp Center 64 SP

All experiments of this paper are conducted using Python version 3.11.4 on a computer featuring a 12th Gen Intel Core i5 processor and 24 GB RAM.

4.2 Prediction process

Both methods share a common ML process, comprising four stages: (1) data preparation, (2) construction of the ML model, (3) validation of its performance, and (4) deployment of the model [1].

4.3 Data preparation

Preparing data stands as a crucial step in constructing reliable and high-performing ML pipelines, given that the quality of the data significantly influences outcomes. This step involves the transformation of raw data into a refined dataset before employing ML algorithms [1]. As per the illustration in Figure 3, data preparation encompasses five distinct stages.

Figure 3: Data preparation process [1]

Data extraction: The original dataset was extracted from the Enterprise Asset Management (EAM) database, focusing on maintenance activities related to wire cutting machines. This dataset contains historical data of executed maintenance actions, including scheduling starting dates, descriptions of failure types, and the duration required to resolve malfunctions (Time to Repair- TTR).

Machines selection: The number of maintenance actions varies, with some machines having over 700 recorded interventions, while others have fewer than 200. Such unbalanced datasets that contain a larger proportion of samples in one class, leading to classifiers showing high accuracy with the majority class but poor performance in predicting minority classes [34]. Most classification algorithms prioritize reducing error ratios, resulting in various techniques, notably sampling methods, to tackle imbalanced datasets [34]. Therefore, we have selected three Schleuniger CC64 SP machines produced in 2006, 2015 and 2021, as illustrated in Table 3. These machines have been chosen for their relevance and representation within our maintenance scope and balances datasets.

Table 3: Manufacturing data of machines sample

Machine ID	Manufacturing Data
M_1	2006
M_2	2015
M_3	2021

After consolidating all the data collected from various machines, we have created a central database containing 9,229 rows and 15 columns. The earliest recorded maintenance event dates back to June 13, 2021, while the most recent event was logged on June 1, 2023.

Data cleaning: The data cleaning phase constituted a crucial step aimed at enhancing the integrity of our dataset, ensuring its validity, and paving the way for precise results in subsequent analyses. The process unfolded systematically, initiating with the filtration of data rows to retain only those pertinent to our investigation. Consequently, entries associated with machines not manufactured by Schleuniger or machines designed for purposes other than cutting were systematically excluded. Subsequently, non-relevant columns were excised, particularly those featuring a static value (e.g., “manufacturer” = Schleuniger), predominantly null values (e.g., “comment”), or deemed extraneous (e.g., “assigned by,” “assigned to,” etc.). Following this, the Z-score method was applied to identify and eliminate outliers within the TTR and TBF columns. This meticulous step aimed to ensure the accuracy of MTBF and MTTR calculations for each machine. To address the remaining null values, a judicious approach was adopted. Missing values in crucial columns, such as TTR, were imputed with the mean values specific to each machine. Subsequently, the residual rows containing any remaining nulls were removed. The culmination of this comprehensive data cleaning process yielded a refined dataset, setting the stage for robust and dependable analyses.

Data transformation: The “Precision-Based Maintenance Prediction Approach” (PBMP) introduced four key columns. First, “TBF” (Time Before Failure) meticulously documents the duration between maintenance interventions for each machine. Secondly, “TTR” (Time to Repair) represents the duration taken to restore a malfunctioning system to full functionality after an incident [29]. Additionally, the “MTBF” column

calculates the average time before failure for individual machines (Eq. 1). Complementing these, the fourth column, "MTTR", signifies the average duration to restore a system, equipment, or machinery to full operational capability after a breakdown (Eq. 2).
The "Occurrence-Driven Maintenance Prediction Approach" (ODMP), using the same dataset as the previous PBMP method, expands the dataset by incorporating "count" and Mean Estimated Hours "MEH". These additions indicate the frequency of failures on specific dates for each machine and the average recovery time, respectively.

Feature engineering: The 10 columns delineate crucial aspects of machine operations and maintenance.

Table 4: Manufacturing data of machines sample

Feature	Description
Machine ID	serves as a unique identifier
Sched	a timeline for maintenance activities
Process	operation type
Model	machine model
Production date	manufacturing date of a product or batch.
TBF	time between failures
MTBF	mean time between failures
TTR	anticipated tasks duration
MTTR	mean time to repair
NF	frequency of failures on specific dates for each machine

Target variable of both models:

PBMP TBF (continues value).

ODMP NF (assumes discrete integer values ranging from 0 to 3).

4.4 Modeling and training

For both analyses, we implemented and tested various algorithms as outlined in Section 3. Each algorithm generates a predictive model, which undergoes an iterative refinement process through continuous training (utilizing 80% of the dataset) and subsequent testing (using the remaining 20% of the dataset) until achieving a high level of accuracy and performance. The testing dataset consists of the final quarter of records for each machine in our dataset, treated individually.

Throughout the training process, the algorithms identify patterns, and subsequently, the model is tested with the remaining 20% of the data to predict unknown attributes, as detailed by [1]. For each approach, the performance of the diverse algorithms is assessed using the appropriate evaluation metrics. The ensuing results are presented and discussed in the following subsections.

4.5 Evaluation and validation

To evaluate the behavior and performance of each model, we employed the evaluation metrics presented in section 3 specific for each category.

4.5.1 Evaluation and validation of PBMP approach

In the context of the PBMP approach, we employed metrics such as MSE, MAE, and R2. The outcomes of these metrics are presented in Table 5.

These results reveal a relatively modest accuracy, wherein mean estimated hour values fall within the 0 to 1.5 range for the three machines, as illustrated in Figures 4, 5, and 6.

Table 5: Evaluation of the different regression algorithms

Algorithm	MSE	MAE	R^2
Gradient Boosting Regression	0.404188	0.346364	-0.066360
Lasso regression	0.393843	0.332226	-0.035377
Linear Regression	0.396700	0.335484	-0.045205
Neural Network Regression	0.399761	0.337669	-0.054564
Random Forest Regression	0.401709	0.344317	-0.059524
Ridge Regression	0.396693	0.335476	-0.045181
Support Vector Regression	0.439302	0.327881	-0.155772

Journal of Theoretical and Applied Information Technology
29th February 2024. Vol.102. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

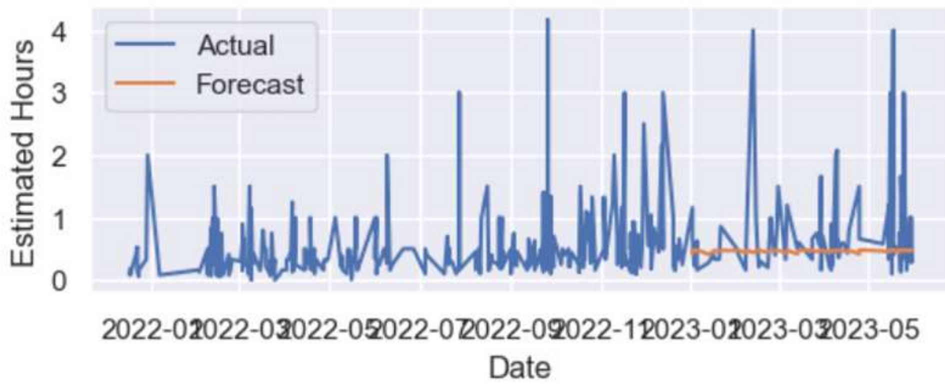

Figure 4: Actual vs Forecast estimated hours of machine ID: 1

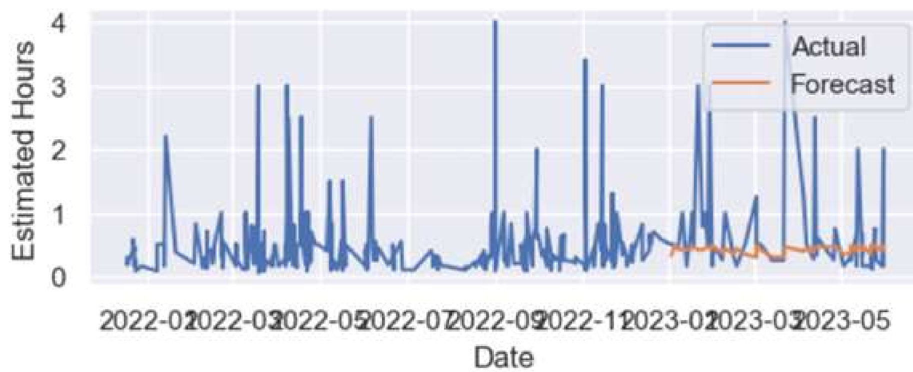

Figure 5: Actual vs Forecast estimated hours of machine ID: 2

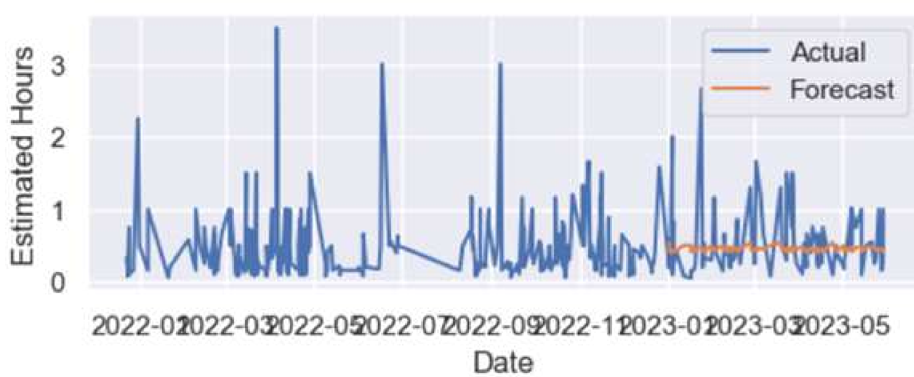

Figure 6: Actual vs Forecast estimated hours of machine ID: 3

4.5.2 ODMP Approach: Classification analysis

For classification analyses the F1 score evaluation metric is chosen as it encompasses both precision and recall values. The outcomes of these analyses for various classification models are detailed in Table 6.

The accuracy of each model, as delineated in Table 6, indicates that, on average, the most effective classification algorithms for predicting machine failure count are Random Forest, followed by LGBM and Decision Tree. These algorithms attained average cross-validation accuracy scores of 71.5%, 67.1%, and 66.4%, respectively, along with average cross-validation F1 scores of 0.398, 0.374, and 0.374. Conversely, the remaining algorithms demonstrated lower accuracy levels.

Confusion matrices are also employed to visually depict the outcomes of the classification models

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and assess the alignment between predicted number of failures and the actual number of failures within the testing dataset (refer to Figures 7, 8, and 9).

Table 6: F1 score and Accuracy of the different classification algorithms

Algorithm	F1_score	Accuracy
Random Forest	0,398	71,50%
KNN	0,292	57,00%
Decision Tree	0,374	66,40%
LGBM	0,374	67,10%

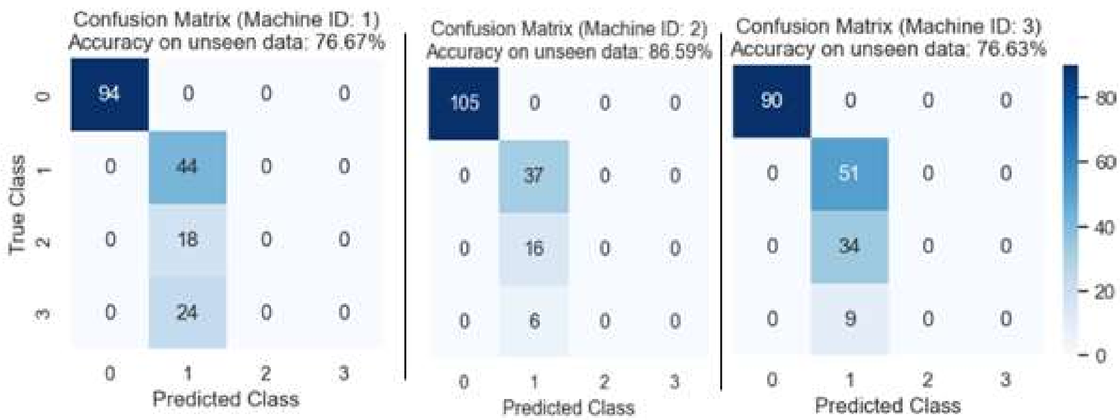

Figure 7: Confusion matrices for machines 1, 2 & 3 generated using Random Forest Algorithm.

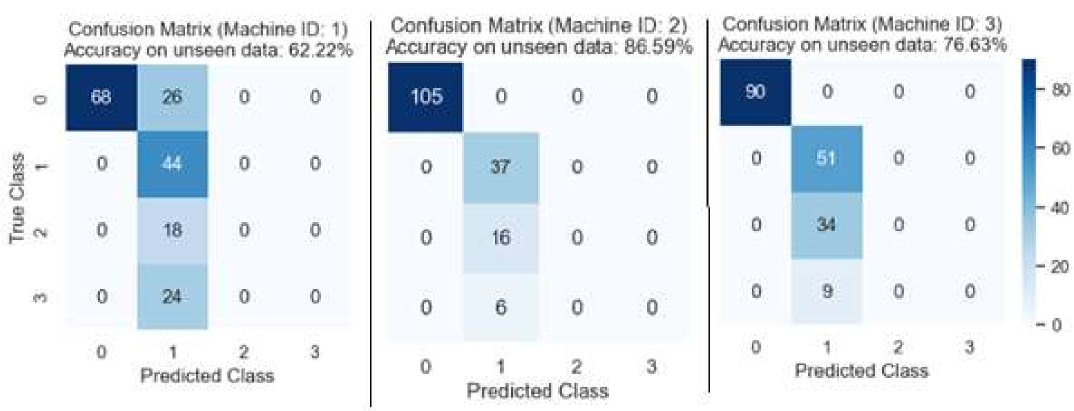

Figure 8: Confusion matrices for machines 1, 2 & 3 generated using LGBM Algorithm.

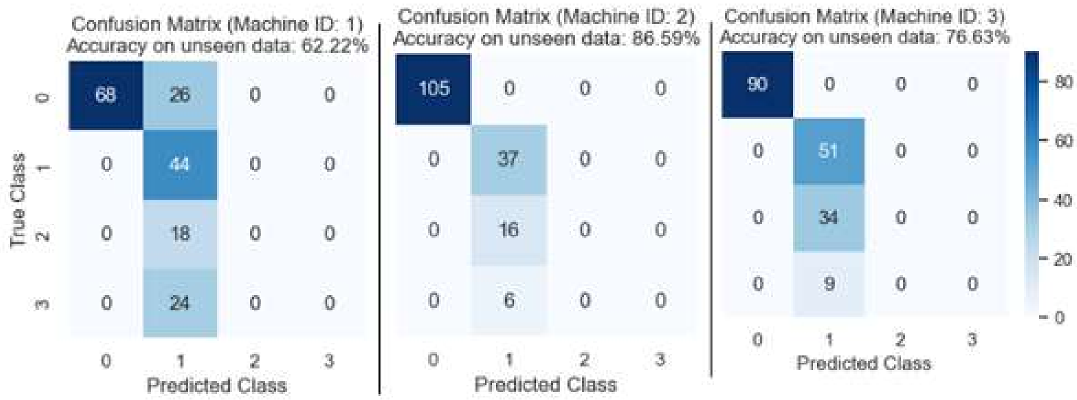

Figure 9: Confusion matrices for machines 1, 2 & 3 generated using Decision Tree Algorithm.

Subsequently, on a one-to-one classification basis, each machine appears to exhibit superior accuracy when utilizing the random forest classifier. Conversely, the results obtained using the LGBM and Decision Tree classifiers seem relatively comparable for every machine.

4.6 Deployment

Leveraging the accomplishments of our training dataset, we opted to implement the random forest model, given its noteworthy performance in prior phases. Figure 10 illustrates the outcomes of our predictive endeavors for the months spanning June to December 2023 for machine ID 1.

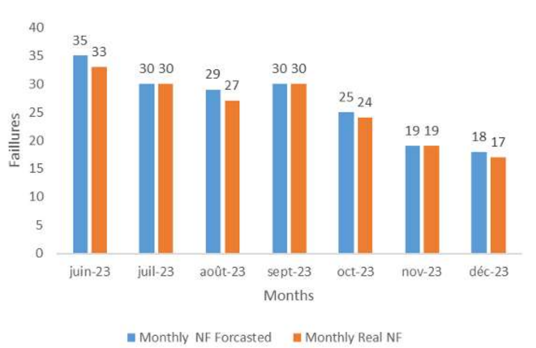

Figure 10: Monthly evaluation: Actual vs Forecast number of failures

Upon examining the graphical representation of the monthly evolution of actual versus predicted numbers of failures (Figure 11), it becomes evident that our model exhibits a high degree of predictive accuracy.

Our emphasis on real-time decision-making and long-term performance evaluation aligns our approach with industry best practices, fostering continuous improvement and precision in predictive maintenance. In fact, the ability to accurately predict the number of failures allows us to proactively anticipate the required number of maintainers, thereby reducing total downtime. This capability not only improves operational efficiency but also contributes to significant cost savings by minimizing disruptions and the need for reactive maintenance. Additionally, the methodology aids in production planning, reduces costs associated with inventory, and exerts a positive impact on the environment by minimizing waste and energy consumption linked to emergency repairs.

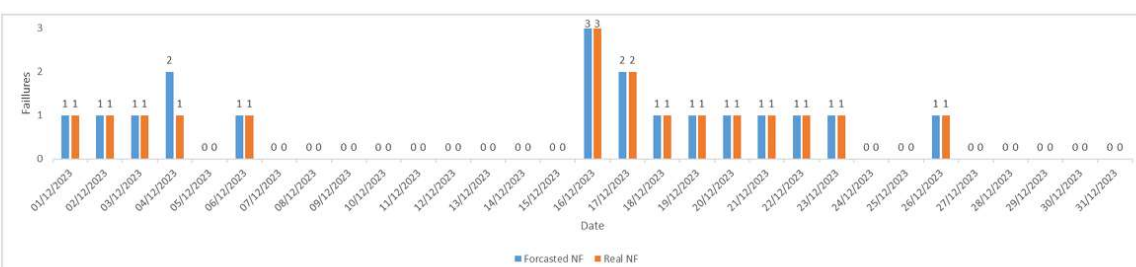

Figure 11: December: Actual vs Forecast number of failures

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION

The evaluation of the two proposed predictive maintenance approaches, utilizing a comprehensive dataset from Schleuniger cutting machines, elucidates the potential of extracting valuable insights from historical maintenance data for proactive decision-making. The results consistently underscore the superiority of the approach focused on predicting the frequency of failure occurrences. While the results also underscore that the considered KPIs may not be sufficient to predict the remaining useful life of the machine. In this context, the incorporation of additional features becomes imperative for more precise predictions.

The deployment of our methodology offers several advantages. Firstly, leveraging insights from previous modeling endeavors enhances the accuracy of predictions. The data-driven approach, coupled with a proactive maintenance strategy, ensures timely responses to potential issues, thereby bolstering machine availability. The deployment of the random forest model, renowned for its higher performance, further solidifies the reliability of our predictions.

This study has opened new perspectives for future research. With access to larger datasets, the refinement of failure predictions to accurately forecast precise failure times, down to the hour or minute, becomes a viable objective. Additionally, expanding predictive capabilities to encompass specific types of failures provides a more comprehensive understanding of issues, potentially expediting recovery times. The utilization of deep learning algorithms on larger datasets and the integration of the categorical boosting method with meta-heuristic algorithms are identified as potential strategies to further enhance predictive capabilities. This forward-looking perspective aims to advance the field of predictive maintenance, providing more nuanced and accurate insights for optimal operational planning and resource allocation in manufacturing systems.

REFERENCES

[1] C. Aurelien, B. Vinnie, P. Rachel, and K. Fawad. “Machine Learning with Dynamics 365 and Power Platform: The Ultimate Guide to Apply Predictive Analytics”. John Wiley, 2022.

[2] O. Ayadi, I. El-Hassani, N. Barka, and T. Masrour. “Real-time kpi forecasting with 1d convolu tional time series for enhanced manufacturing efficiency”. *In International Conference on Artificial Intelligence & Industrial Applications*, pages 36–47. Springer, 2023.

[3] S. Behera and R. Misra. “Generative adversarial networks based remaining useful life estimation for iiot”. *Computers & Electrical Engineering*, 92:107195, 2021

[4] G. Bencheikh, A. Letouzey, and X. Desforges. “An approach for joint scheduling of production and predictive maintenance activities”. *Journal of Manufacturing Systems*, 64:546–560, 2022.

[5] [5] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak. “Health assessment and life prediction of cutting tools based on support vector regression”. *Journal of intelligent manufacturing*, 26:213–223, 2015.

[6] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcala. “A systematic literature review of machine learning methods applied to predictive maintenance”. *Computers & Industrial Engineering*, 137:106024, 2019.

[7] E. G. Gorski, E. D. F. R. Loures, E. A. P. Santos, R. E. Kondo, and G. R. D. N. Martins. “Towards a smart workflow in cmms/eam systems: An approach based on ml and mcdm”. *Journal of Industrial Information Integration*, 26:100278, 2022.

[8] G. Hesamian, A. Johannssen, and N. Chukhrova. “An explainable fused lasso regression model for handling high-dimensional fuzzy data”. *Journal of Computational and Applied Mathematics*, 441:115721, 2024.

[9] Y. Huang, Y. Lei, X. Luo, and C. Fu. “Prediction of compressive strength of rice husk ash concrete: a comparison of different metaheuristic algorithms for optimizing support vector regression”. *Case Studies in Construction Materials*, page e02201, 2023.

[10] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne. “Artificial intelligence and real time predictive maintenance in industry 4.0: a bibliometric analysis”. *AI and Ethics*, 2(4):553–577, 2022.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[11] Y. Kong and K. Kurumisawa. "Application of machine learning in predicting workability for alkali activated materials". *Case Studies in Construction Materials*, page e02173, 2023.

[12] S. Maataoui, G. Bencheikh, and G. Bencheikh. "Predictive maintenance in the industrial sector: A crisp-dm approach for developing accurate machine failure prediction models". *In 2023 Fifth Inter national Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, pages 223–227. IEEE, 2023.

[13] T. Mahakul, S. Baboo, S. Patnaik, and F. UNIVERSTY. "Implementation of enterprise asset management using it tools: A case study of ib thermal power station". *Journal of Information Technology Management*, 16(4):39, 2005.

[14] M. C. May, Z. Fang, M. B. Eitel, N. Stricker, D. Ghoshdastidar, and G. Lanza. "Graph-based prediction of missing kpis through optimization and random forests for kpi systems". *Production Engineering*, 17(2):211–222, 2023.

[15] G. C. McDonald. "Ridge regression". *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009.

[16] K. Medjaher, N. Zerhouni, and R. Gouriveau. "From prognostics and health systems management to predictive maintenance 1: Monitoring and prognostics". John Wiley & Sons, 2016.

[17] K. T. Nguyen, K. Medjaher, and D. T. Tran. "A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines". *Artificial Intelligence Review*, 56(4):3659–3709, 2023.

[18] H. NR. "Performance analysis of spinning machines using machine learning". *In Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021)*, 2021.

[19] S. Ochella, M. Shafiee, and F. Dinmohammadi. "Artificial intelligence in prognostics and health management of engineering systems". *Engineering Applications of Artificial Intelligence*, 108:104552, 2022.

[20] M. Orosnjak and D. Sevic. "Benchmarking maintenance practices for allocating features affecting hydraulic system maintenance: A west-balkan perspective". *Mathematics*, 11(18):3816, 2023.

[21] N. Rahpeymai. "Data Mining with Decision Trees in the Gene Logic Database: A Breast Cancer Study". Institutionen for datavetenskap, 2002.

[22] S. M. Rezvanizaniani, J. Dempsey, and J. Lee. "An effective predictive maintenance approach based on historical maintenance data using a probabilistic risk assessment: Phm14 data challenge". *International Journal of Prognostics and Health Management*, 5(2), 2014.

[23] S. J. Rigatti. "Random forest". *Journal of Insurance Medicine*, 47(1):31–39, 2017.

[24] M. G. Rojas, A. C. Olivera, and P. J. Vidal. "A genetic operators-based ant lion optimiser for training a medical multi-layer perceptron". *Applied Soft Computing*, page 111192, 2023.

[25] J. R. Rokde and A. G. Thosar. "Linear regression approach for performance evaluation of es with load impedance variations of non-critical and critical load". *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 6:100312, 2023.

[26] V. Samsonov, E. Chrismarie, H.-G. K¨opken, S. B¨ar, D. L¨ utticke, and T. Meisen. "Deep repre sentation learning and reinforcement learning for workpiece setup optimization in cnc milling". *Production Engineering*, 17(6):847–859, 2023.

[27] M. R. Segal. "Machine learning benchmarks and random forest regression". 2004.

[28] D. F. Specht. "A general regression neural network". *IEEE transactions on neural networks*, 2(6):568–576, 1991.

[29] V. G. Talkit, P. Honrao, S. Bhosale, A. Bewoor, and R. B. Patil. "Application of the reliability analysis for modifications in maintenance scheduling". *Materials Today: Proceedings*, 77:579–585, 2023.

[30] X. Wang, X. Wang, B. Ma, Q. Li, C. Wang, and Y. Shi. "High-performance reversible data hiding based on ridge regression prediction algorithm". *Signal Processing*, 204:108818, 2023.

[31] L. Xiong and Y. Yao. "Study on an adaptive thermal comfort model with k-nearest-neighbors (knn) algorithm". *Building and Environment*, 202:108026, 2021.

[32] N. Xu, Z. Wang, Y. Dai, Q. Li, W. Zhu, R. Wang, and R. B. Finkelman. "Prediction of higher heating value of coal based on gradient boosting regression tree model". *International Journal of Coal Geology*, page 104293, 2023.

[33] F. Yavari and N. Pilevari. “Industry revolutions development from industry 1.0 to industry 5.0 in manufacturing”. *Journal of Industrial Strategic Management*, 5(2):44–63, 2020.

[34] P. Yıldırım. “Pattern classification with imbalanced and multiclass data for the prediction of albendazole adverse event outcomes”. *Procedia Computer Science*, 83:1013–1018, 2016.

[35] S. Zhai, M. G. Kandemir, and G. Reinhart. “Predictive maintenance integrated production schedul ing by applying deep generative prognostics models: approach, formulation and solution”. *Produc tion Engineering*, pages 1–24, 2022.

[36] T. Zonta, C. A. Da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li. “Predictive maintenance in the industry 4.0: A systematic literature review”. *Computers & Industrial Engineering*, 150:106889, 2020.