applied sciences

MDPI

Article

Predictors of Successful Maintenance Practices in Companies Using Fluid Power Systems: A Model-Agnostic Interpretation

Marko Orošnjak *, Ivan Beker, Nebojša Brkljač and Vijoleta Vrhovac

Department of Industrial Engineering and Engineering Management, Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia; beker@uns.ac.rs (I.B.); n.brkljac@uns.ac.rs (N.B.); violeta.vrhovac@uns.ac.rs (V.V.)

* Correspondence: orosnjak@uns.ac.rs

Abstract: The study identifies critical factors influencing companies' operational and sustainability performance utilising fluid power systems. Firstly, the study performs Machine Learning (ML) modelling using variables extracted from survey instruments in the West Balkan region. The dataset comprises 115 companies (38.75% response rate). The survey data consist of 22 predictors, including meta-data and three target variables. The K-Nearest Neighbours algorithm offers the highest predictive accuracy compared to the other seven ML models, including Ridge Regression, Support Vector Regression, and ElasticNet Regression. Next, using a model-agnostic interpretation, we assess feature importance using mean dropout loss. After extracting the most essential features, we test hypotheses to understand individual variables' local and global interpretation of maintenance performance metrics. The findings suggest that Failure Analysis Personnel, data analytics, and the usage of advanced technological solutions significantly impact the availability and sustainability of these systems.

Keywords: machine learning; feature extraction; fluid power systems; multivariate statistics

check for updates

Citation: Orošnjak, M.; Beker, I.; Brkljač, N.; Vrhovac, V. Predictors of Successful Maintenance Practices in Companies Using Fluid Power Systems: A Model-Agnostic Interpretation. *Appl. Sci.* **2024**, *14*, 5921. https://doi.org/10.3390/app14135921

Academic Editor: Jose Machado

Received: 29 May 2024
Revised: 26 June 2024
Accepted: 5 July 2024
Published: 6 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Overall Equipment Effectiveness (OEE) is no longer solely a target of manufacturing companies. Many companies are altering their business models and incorporating sustainability aspects in their performance metrics [1–4]. Specifically, in asset-intensive industries where processes and production heavily rely on fluid power systems, ensuring the availability and sustainability of these assets is crucial.

Recent advancements in digital technology offer new opportunities for advancing maintenance practices. As such, an exponential rise in publications dedicated to Predictive Maintenance (PdM), Energy-Based Maintenance (EBM), and Sustainable Maintenance (SM) has been reported [5–9]. However, adopting these advanced maintenance practices varies by company and is questionable in terms of suitability and readiness. This is especially important due to the heterogeneity of applications relying on fluid power systems and other factors, such as the availability of skilled personnel [10].

Due to limited knowledge and lack of empirical evidence, the study aims to investigate critical factors affecting the operational and sustainability performance of companies utilising fluid power systems. To achieve this, we rely on evidence from a questionnaire-based survey [11] of companies in the West Balkan region. The idea is to raise awareness and understand how specific maintenance activities, tools, and organisational characteristics affect the performance metrics expressed in Mean Time Between Failures (MTBF), Mean Time to Repair (MTTR), and Waste Oil Monthly (WOM).

There is a lack of knowledge in contemporary maintenance research regarding the factors impacting the maintenance performance metrics [12]. With the emergence of SM, many studies engage in observational and explanatory research relying on survey data. For instance, a study by [13] investigates the impact of traditional preventive maintenance

versus EBM practices, suggesting that although EBM may offer benefits in reducing energy consumption and environmental impact, it significantly reduces operational effectiveness in terms of MTBF, even up to 53% drop due to lack of competences and skills in EBM domain. A recent study by Judijano et al. [14] using SEM (Structural Equation Modelling) highlighted the impact of business analytics and big data on predictive maintenance and asset management practices, suggesting a significant positive relationship with the use of data analytic tools. The findings suggest significant improvements in asset reliability, reduced downtime, and enhanced operational efficiency. From an operational perspective, the role of management commitment and TPM (Total Productive Maintenance) implementation strategies suggest a significant benefit to companies' productivity, as reported by Diaz-Reza et al. [15]. They conclude that managerial commitment, as a latent variable, accounts for 40.7% of the variation, implying that for a maintenance program to be successful, managerial support is crucial for increasing maintenance performance. Their findings demonstrate with a moderate to large effect size that management commitment is vital in productivity, TPM success, and PM programs. Lastly, considering the contextual settings regarding the impact of factors on companies relying on fluid power systems, a study conducted by Orošnjak et al. [16] using a regression model suggests that filter replacement time, machine age, Failure Analysis Personnel, and maintenance practice significantly impact MTBF ($p < 0.01$).

The study continues previous work [11] to assess the factors influencing the maintenance performance of assets using fluid power systems. However, instead of investigating the factors leading to specific component failures, failure modes, and root causes of failure, this study relies on quantitative analysis using the Model-Agnostic Interpretation (MAI) method [17] of utilised Machine Learning (ML) models. The rationale for selecting a specific methodology is due to the underlying maintenance features' complex and nonlinear relationship. Specifically, we aim to identify the best-performing models and offer decision-makers better insight into the features' contribution. The MAI method helps understand the versatility across models by relying on the MDL (Mean Dropout Loss) algorithm. The MDL is not tied to a specific model but provides feature importance based on the loss function. In contrast to the traditional approach of simply relying on the model's performance and using features of the best-performing model, leading to biased estimates and ignoring essential features, we offer the quantitative impact of individual features and identify critical factors affecting maintenance performance. Such an approach compares features across models and aids in understanding feature contribution. In sum, the approach is considered more robust than solely relying on ML models' metrics that cannot be compared across models. Lastly, after extracting the most relevant features, we conduct a robust statistical analysis of individual features using a variety of hypothesis testing for categorical and continuous predictors.

The rest of the manuscript is structured as follows. The second section describes the questionnaire-based survey, including the description of retrieved samples and features used in modelling. The section explains data (pre)processing, data wrangling, ML models with settings of (hyper)parameters, and the MDL algorithm. The third section offers in-detail results of descriptive statistics, ML performance metrics, and hypothesis testing findings derived from parametric and non-parametric analyses. The last section provides concluding remarks, as well as the limitations and implications of the study.

2. Materials and Methods

2.1. Survey Instrument and Data

The questionnaire-based survey instrument has been developed to target manufacturing companies in the West Balkan region. The survey consists of eighteen predictor variables, four meta-data records, and five target variables (Figure 1). The survey's response rate is 38.72% (115/297). The extracted variables were subjected to coding. Specifically, NWEC (Nominal Working Energy Consumption) is determined based on the NWP (Nominal Working Pressure) and NWF (Nominal Working Flow) metrics, expressed in kW power.

Similarly, the Maintenance Personnel Per Machine (MPPM) was determined based on the number of Maintenance Personnel (nMANP) and Number of Machines (NOM) utilising fluid power systems. The rest of the survey data are taken as raw data. The full description and design of the survey instrument is provided in detail [11]. However, unlike in the previous study, we specifically aim to assess the impact of individual predictors on the MTTR, MTBF, and WOM metrics. Thus, RCF (Root Cause of Failure) and CFT (Component Failure Type) variables are not included in this study.

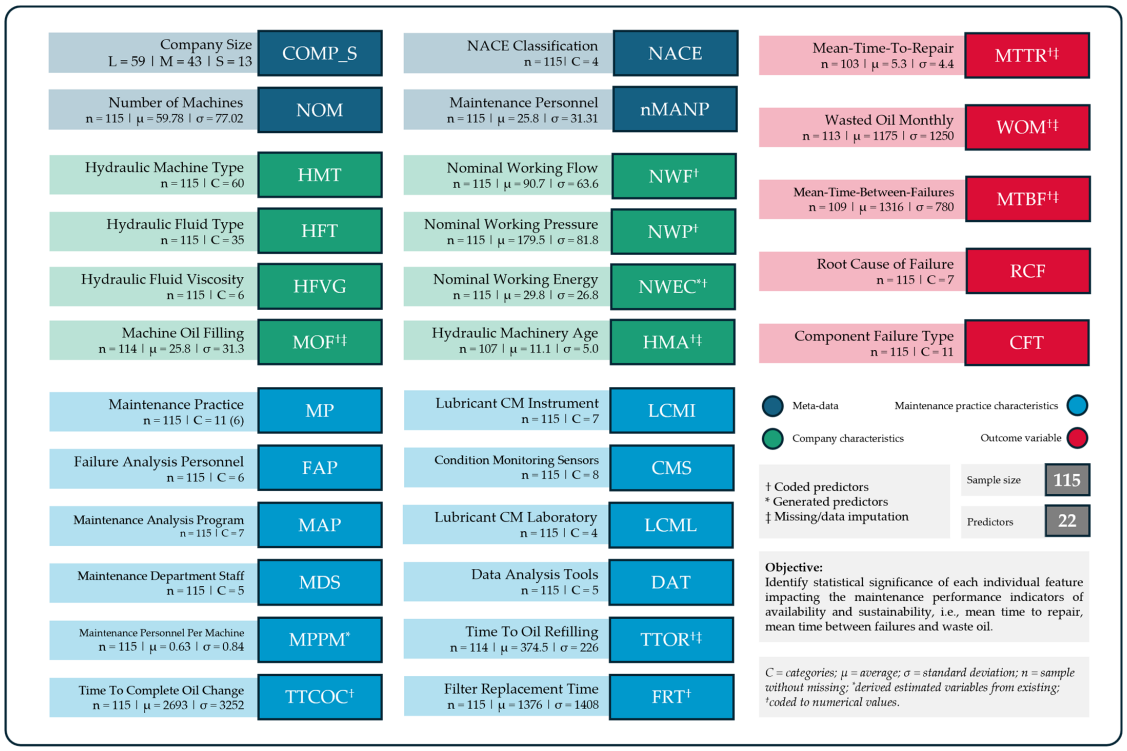


Figure 1. The questionnaire-based survey description of input and output variables and estimated parameters (reworked [17]).

Moreover, additional coding of variables is performed. For FAP (Failure Analysis Personnel), the data are split into indoor and outdoor (outsource) Failure Analysis Personnel because we consider it would increase the generalisability of findings that would be more of an interest to the readers. Also, the LCML (Lubricant Condition Monitoring Laboratory) variable is coded as a binary variable as 1 = includes laboratory analysis (e.g., oil spectrophotometry) and 0 = does not include laboratory analysis. The LCMI (Lubricant Condition Monitoring Instruments) variable is also coded as binary, such that 1 = includes advanced condition monitoring instruments (e.g., contamination sensors and vibroacoustic sensors), and 0 = does not include. Lastly, the DAT (Data Analytics Tools) utilised variable is coded as 1 = includes data analysis tools and 0 = does not include data analysis tools.

The rest of the categorical and continuous predictors are explained and discussed through hypothesis testing, including the impact of each category and the association between continuous and target variables. Lastly, MTTR, MTBF, and WOM's output performance metrics were subjected to dimensionality reduction. Specifically, we use PCA (Principal Component Analysis) and select the first PC (Principal Component), which is commonly used since it captures the most variance. Next, we perform a regression analysis of extracted components using coded categorical and continuous variables to capture the contribution of each feature in the PC1 using the MDL algorithm. The results of dimensionality reduction using PCA are provided below.

The PCA is a commonly known unsupervised method used for dimensionality reduction and visualisation. [18]. Namely, the PCA transforms the original dataset into a new set

of variables through ordination, commonly known as Principal Components [19]. These PCs are then ordered by the variation they capture from the original dataset. The reader is referred to for an in-detail analytical description of the PCA [20,21]. The dimensionality reduction results using PCA show the loadings (Table 1) of individual components and PC coefficients (Table 2). The variance for the first two components accounts for 79.7% of the total variance, while the projected first component accounts for 48.8% of the total variance. The PC1 is then used to perform an ML regression analysis using the standardised raw data, after which the contribution of individual predictors is determined by the MDL. The ML regression models are explained in the following.

Table 1. PCA loadings.

Variable	PC1	PC2	PC3	Cumulative
MTBF_St	−0.7990	−0.2305	0.5554	1.000
MTTR_St	0.4445	−0.8955	0.0224	1.000
WOM_St	0.7917	0.2702	0.5480	1.000
%Variance	48.8%	30.9%	20.3%	100%

Table 2. Factor coefficients.

Variable	PC1	PC2	PC3
MTBF	−0.5463	−0.2483	0.9116
MTTR	0.3039	−0.9649	0.0367
WOM	0.5412	0.2911	0.8995

2.2. Machine Learning Models

2.2.1. Multiple Linear Regression

The model is developed as a simple linear regression model. Although it is a commonly used multivariate statistical technique, it is also one of the most frequently used ML tools in fault detection and diagnosis [22–24]. Given the extent of our analysis, the analytical formulation of linear regression and other ML techniques are left out and described elsewhere [25,26]. However, to ensure the replicability of our findings, we provide a complete description of the parameters for every ML model. All models and results can be replicated using settings in the open-source software JASP (v.0.18.3).

2.2.2. Gradient Boosting Regression

The algorithmic settings consist of shrinkage = 0.1, interaction depth = 1.0, minimum observations per node = 10, training data used per tree = 50%, and loss function = t (since Gaussian and Laplace show lower model performance on the holdout dataset). The number of trees is determined based on optimisations of 100 trees. The features are scaled based on standardisation, and the seed is set to 2805. The sample size is set to split 80/20, while training and validation are performed using a 10-fold cross-validation procedure.

2.2.3. k-Nearest Neighbours Regression

It is not a commonly used ML regression model; however, we decided to include the kNN model as, in some cases, it reports good regression performance [27–29]. The model parameters are set: distance = Euclidian; the selected k number of nearest neighbours is determined based on the optimisation with a maximum of k = 10 nearest neighbours. The weights are set to inverse. Given that a broad spectrum of weights can be utilised for the kNN algorithm (e.g., rectangular, Gaussian, Epanechnikov, and cosine) [30], we performed the analysis on ten proposed weights reported in the literature [31–33], and the model shows the highest performance using inverse weights.

2.2.4. Artificial Neural Network Regression

The settings for training and testing the ANN included the following. We first tested the performance using different settings of the ANN algorithm. The performance selection between BP (Back Propagation) and RPROP (Resilient Back Propagation) variants [34]. Included RPROP variants are RPROP+ (Resilient Propagation with Backtracking), RPROR- (Resilient Propagation without Backtracking), GLPORP-SAG (Globally Convergent Resilient Propagation with Smallest Absolute Gradient), and GLPORP-SLR (Globally Convergent Resilient Propagation with Smallest Learning Rate) [35]. The stopping criteria loss function is set to 1.0 with maximum training repetitions = 10^5.

For setting the network topology, we did not subjectively (manually) set the network. Instead, we optimised using GA (Genetic Algorithm) [36–39]. The hyperparameters of GA are set as population size = 20; number of generations = 10; and the number of hidden layers is set to a maximum of 10 layers with a maximum number of 10 nodes in each hidden layer of the network. The parent selection is set to roulette wheel (we performed different combinations of parent selections using universal, ranked, tournament, random, and roulette wheel showed the highest performance); the uniform crossover method is selected (one-point and multipoint crossover method was also tested and uniform crossover method showed highest performance); mutations are set to reset (inversion, swap, and scramble were tested and reduced the performance) with 10% probability; and the fitness-based is set as a survival method (age-based and random survival were tested and resulted in poorer performance) with 10% elitism.

2.2.5. Regularised Linear Regression

Regularised linear regression is performed using different algorithm settings. First, we performed regularisation using the Lasso penalty function, followed by the Ridge and ElasticNet functions. The intercept is included, and lambda is optimised.

2.2.6. Support Vector Regression

The Support Vector Regression (SVR) is used with different algorithm settings. First, we set weights to radial–linear, polynomial, and sigmoid, which reduce the model's performance significantly. The gamma parameter of the radial function is set to 1.0. The tolerance of termination criteria is set to 0.001, while parameter ξ = 0.01. The costs of constraint violations are not fixed but instead are used with optimisation, with the maximum violation cost set to 5. The features are scaled with the seed set to 2805.

2.2.7. Performance Metrics

The common practice of evaluating or quantifying the errors between predicted and actual results is used to measure the performance of the applied model. Thus, the following performance metrics are used. MSE (Mean Squared Error):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{1}$$

where y_i is the actual value, $\hat{y}_i$ is predicted value, and n is the observation count. The RMSE (Root Mean Squared Error) is estimated as the following:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{2}$$

The MAE (Mean Absolute Error)/Mean Absolute Deviation (MAD) is determined the following:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{3}$$

The MAPE (Mean Absolute Percentage Error) is determined as the following:

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \quad (4)$$

The coefficient of approximation R^2 is determined as the following:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}. \quad (5)$$

where $\overline{y}_i$ is the mean of the actual values.

2.3. Feature Importance and Mean Dropout Loss

The last step of the analysis is determining the variable, i.e., feature importance from the proposed algorithms. For estimating the relative influence of predictor variables, we rely on MDL (Mean Dropout Loss) [40–42]. The MDL is a commonly used metric in ML for estimating the relative influence that individual variable (s) have on model performance. For understanding, we first explain the objects used in the Algorithm 1. The X represents a matrix of input variables (features) with dimensions of $n \times$ p, whereas the n is the number of observations, and p is the number of variables in the processed dataset. The y is the output vector of dimension n, which in this case, represents the component PC1. The $\mathcal{M}$ is the ML model, and $\mathcal{L}$ is the loss function, in this case, MSE. The algorithm output is feature importance point value $\mathcal{F}$.

Algorithm 1. Determining the feature importance based on Mean Dropout Loss.

Input X, y, $\mathcal{M}$, $\mathcal{L}$
Output $\mathcal{F}$ feature importance
Steps
1 Train the baseline model $\mathcal{M}$
Model $\mathcal{M}$ training using training set (X_{train}, y_{train})
Perfom prediction on a holdout dataset X_{hold} for obtaining $\mathcal{M}(X_{hold})$
Calculate baseline loss function $\mathcal{L}$: $\mathcal{L}_{\text{baseline}} = \mathcal{L}(y_{hold}, y_{baseline})$
2 Initilise feature importance $\mathcal{F}$
Let $\mathcal{F}$ be a vector with length p such that $\mathcal{F}$ = [0, 0 . . ., 0]
3 Calculate *MDL* **for each feature**
for each feature j in {1, 2, 3,. . . p}
create validation on a holdout set $X_{hold}^{(j)}$ where j-th feature is dropped after 50 permutation set with its mean value such that:
$X_{hold}^{(j)} = X_{hold}$ with $X_{hold;y}$ dropped
perform prediction on $X_{hold}^{(j)}$ to obtain $\hat{y}^{(j)}$:
$\hat{y}^{(j)} = \mathcal{M}\left(X_{hold}^{(j)}\right)$
calculate the loss $\mathcal{L}^{(j)}$:
$\mathcal{L}^{(j)} = \mathcal{L}\left(y_{hold}, \hat{y}^{(j)}\right)$
perform *MDL* for the j-th feature:
$MDL_j = \mathcal{L}^{(j)} - \mathcal{L}_{baseline}$
assign value for MDL_j based on $\mathcal{F}_j$:
$\mathcal{F}_{[j]} = MDL_j$
end for
4 (optional) Normalisation of feature importance $\mathcal{F}_j$:
normalisation of features $\mathcal{F}_{norm.} = \frac{\mathcal{F}}{\sum_{j=1}^{p}\mathcal{F}_{[j]}}$
5 Return

3. Results and Discussion

We provide descriptive statistics, ML, and inferential statistical results in the following. After performing ML modelling on the training–testing dataset, the extracted features are used for hypothesis testing. These are designed to understand the impact and the amount of variance shared with individual target variables. Understanding the contribution and effect of these predictors will certainly aid maintenance decision-making regarding techno-economic and sustainable benefits.

3.1. Descriptive Statistics

The descriptive analysis includes meta-data and categorical and continuous predictors used in ML regression modelling. Firstly, the size of companies consists of large (57), medium (41), and small (13) companies, classified according to NACE into agriculture, forestry and fishing (10), construction (22), manufacturing (51), and mining and quarrying (28). There are a total of 6746 in Number of Machines (NoM), i.e., fluid power systems identified in the sample with a total of n = 2775 Maintenance Personnel (nMANP), specifically NoM = 60.775 $\pm$ 78.132, and nMANP = 25.00 $\pm$ 29.244.

Considering the HFVG variable, it was difficult to extract a specific category by text mining; however, the most common fluids include viscosity grade VG32, VG46, and VG68, with VG46 being the most commonly used. The most common types of assets and machinery (HMT) utilised include excavators, CNC machines, material handling machines, manipulators, die casting, production lines, hydraulic presses, extruding machines, etc. The most common Hydraulic Fluid (HFT) types are HV 29.14%, HL 19.43% and HM 8.57%, among others. The rest of the asset characteristics are provided in Table 3.

Table 3. Descriptive statistics of assets relying on fluid power transfer or control.

	Mean	95%CI_L	95%CI_U	St. dev.	Min	Max	Sum
MOF_Est	432.869	305.850	570.501	742.147	50.000	3500.000	48,048.460
HMA_Est	10.749	9.933	11.622	4.420	2.500	35.000	1193.170
NWP_Est	178.385	164.393	193.927	81.805	65.000	415.250	19,800.740
NWF_Est	90.165	78.146	102.117	63.118	20.000	242.550	10,008.290
NWEC_Est	29.407	25.032	34.729	26.160	2.170	113.390	3264.170
MPPM	0.611	0.464	0.792	0.838	0.059	5.000	67.837
FRT_Est	1382.40	1149.69	1657.75	1420.89	200.0	8640.00	153,446.49
TTOR_Est	382.11	344.173	423.07	225.76	25.00	1000.00	42,414.13
TTCOC_Est	2718.81	2096.76	3340.86	3307.014	100.00	20,000.00	301,787.90

The maintenance predictors provide the following insights. The types of maintenance practices identified are overlapping where most applied include PM (Preventive Maintenance) in 73.87% of cases, followed by FBM (Failure-Based Maintenance) or Corrective Maintenance in 34.23%, CBM (Condition-Based Maintenance) in 41.44%, PdM (Predictive Maintenance) 10.81%, OM (Opportunity Maintenance) 3.6%, and DM (Design-out Maintenance) 0.9%. The formulation of proposed practices is developed according to [43]. The FAP explains the personnel responsible for performing fault detection and diagnosis, where 49.5% of personnel is outsourced while 50.5% is indoor. The MAP describes the analysis program consisting of companies without any program (8.1%), visual inspections (39.6%), contamination control (12.6%), oil analysis program (22.5%), and full prognostics and health management (17.1%). The CMS predictor describes categories of sensors used to perform condition monitoring—none (19.8%), pressure/flow (21.6%), pressure, flow, and temperature (27.9%), and including those above with additional contamination and vibroacoustic sensors (30.6%). The LCMI is coded as a binary variable explaining the use of specific lubricant condition monitoring instruments (yes—31.5%/no—68.5%). Lastly, the LCML explains the laboratorial analysis performed to support fault detection and diagnosis of the fluid power system (yes—2.7%/no—97.3%). The rest of the continuous predictors (MPPM, FRT, TTOR, and TTCOC) are described in Table 3.

3.2. Machine Learning and Feature Importance

The performance of ML models (Figure 2) suggests the following. The highest performance in terms of coefficient R^2 is obtained for Ridge regression (R^2 = 0.399), SVR (R^2 = 0.38), kNN (R^2 = 0.314) and ElasticNet (R^2 = 0.313). From the perspective of error distributions, the Ridge-RMSE (0.731), ElasticNet-RMSE (0.791), Lasso-RMSE (0.801), and kNN (0.815) have the lowest reported RMSE. However, MSE and RMSE are sensitive to outliers, especially in this study, where a high deviation in the score, even after normalisation, is observed. Thus, MAE/MAD and MAPE may provide much better insight into the model's performance due to lower sensitivity to outliers. The best-performing algorithms include Ridge, ElasticNet, Lasso, and SVR models.

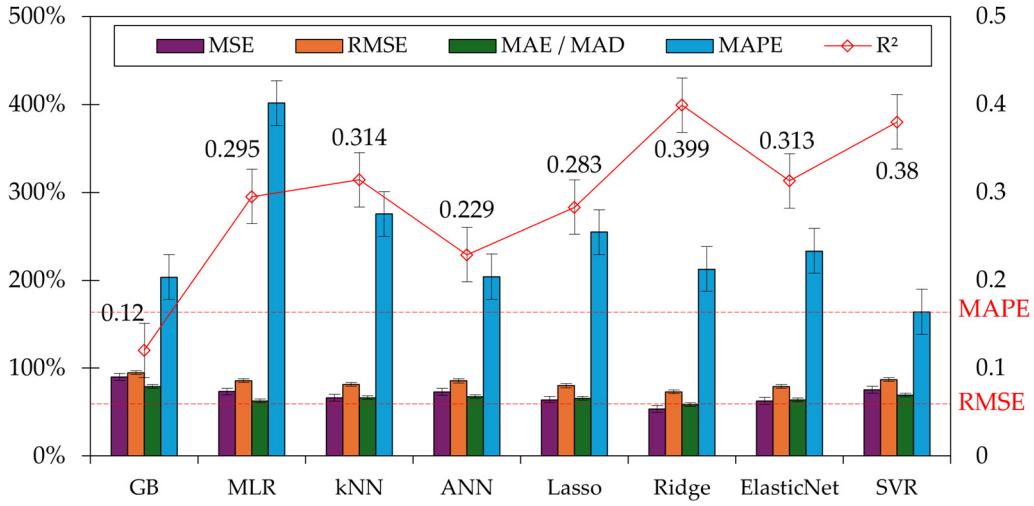


Figure 2. Predictive performance scores.

The predictive performance of samples considering predicted and observed test values (Figure 3) suggests that the least variability is observed in Lasso, Ridge, ElasticNet, and SVR regressions. Specifically, similar outcomes but with smaller point error estimates in elaborating the performance are observed in Ridge regression. It was a bit surprising seeing that even after performing an in-depth design and modification of a Deep Neural Network topology with GA (Genetic Algorithm) using MLP (Multilayer Perceptron) GLPROP-SLR-GA, the maximum regression obtained was R^2 = 0.229. Even so, the complete set of predictors could not be included for a network to converge at the 10^5 maximum number of training repetitions, which forced us to train the network only with continuous variables.

To avoid making biased statements, instead of including or selecting the highest-performing model, we visualise the results of feature importance based on the MDL score for extracting an overall average feature importance score (Figure 4). The overall assessment suggests that FAP_Binary, followed by NWEC_Est, HMA_Est, CMS_Coded, MAP_Coded, TTOR_Est, MDS_Coded, FRT_Est, and MPPM, contribute the most amount of variance to the proposed models. Although the literature often reports Machine Age (HMA) as the most contributing factor to the technical performance metrics, such as MTTR and MTBF, it was surprising that maintenance practice performed is the least contributing factor to the overall ML regression. This forced us to conduct additional hypothesis testing to understand the actual contribution of individual variables (predictors) to the overall ML models by estimating whether or not these predictors have a statistically significant contribution.

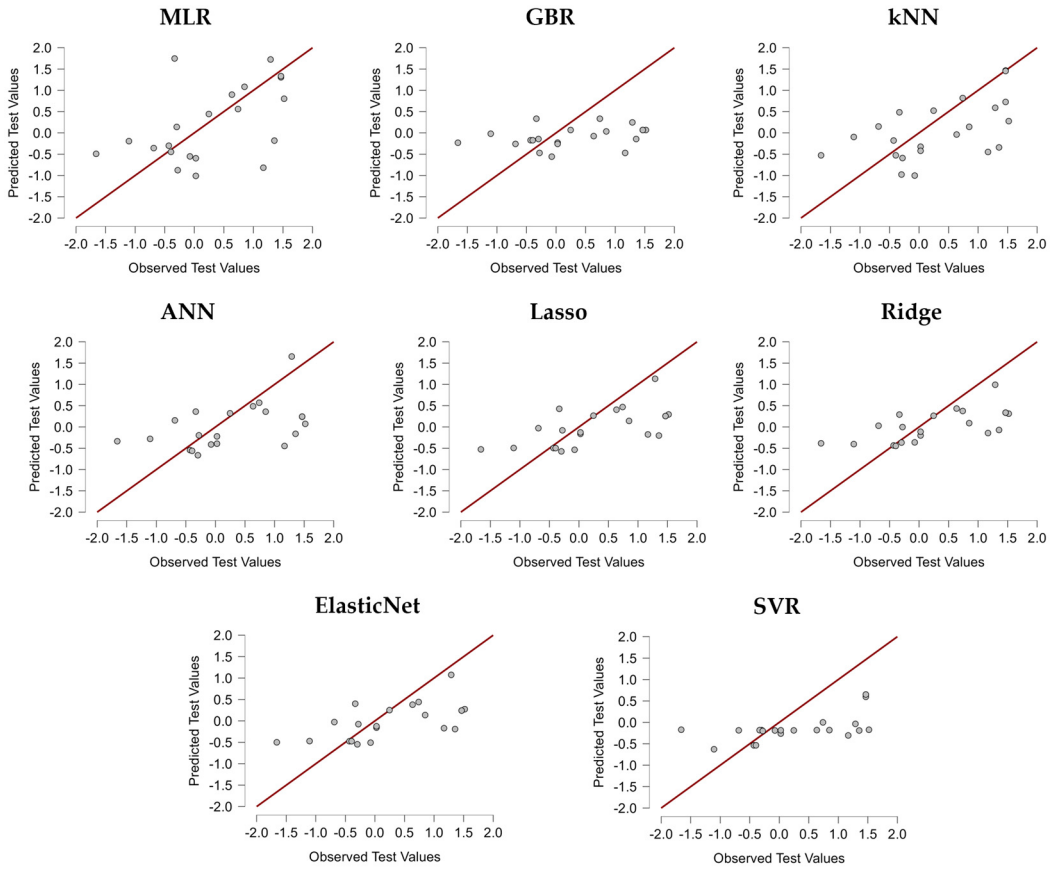


Figure 3. Predictive performance regression plots of the training set.

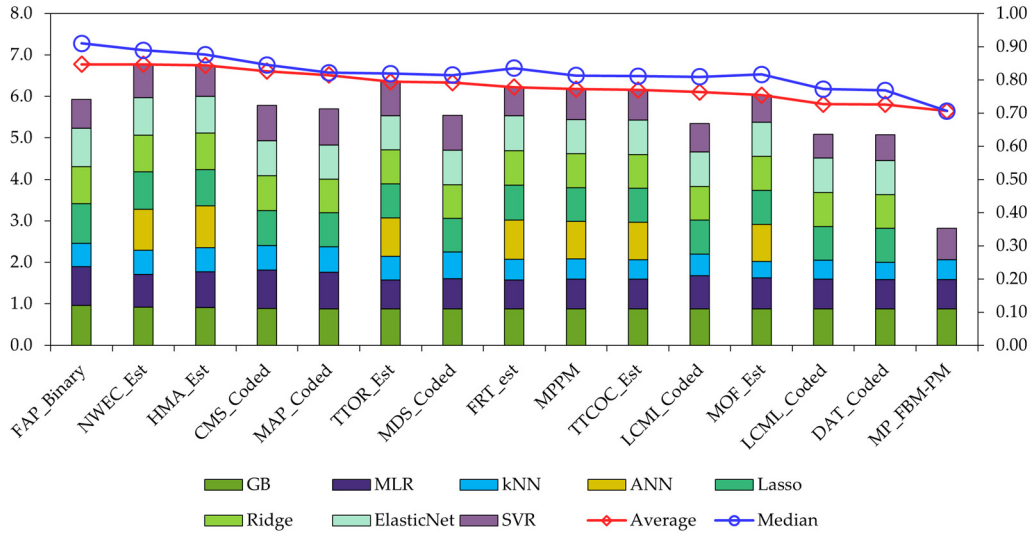


Figure 4. Feature importance score.

3.3. Analysis of Features

Hypothesis tests are performed to understand the impact of predictors on maintenance performance metrics. For binary variables, an independent sample *t*-test statistic is used; for categorical predictors, an ANOVA (Analysis of Variance) test statistic is used; while for continuous variables, a correlation is used to determine linear dependency on target variables. Lastly, the results are reported using parametric and non-parametric statistical tests in cases where test assumptions are violated.

3.3.1. Independent Samples Statistic

Firstly, an analysis of the FAP predictor is performed. The split between indoor and outdoor (outsource) Failure Analysis Personnel is performed. Both parametric and non-parametric results are reported due to possible violation of normality assumptions reported via the Shapiro–Wilk test. For instance, WOM_Est ($W = 0.763$, $p < 0.001$) and MTTR_Est ($W = 0.878$, $p < 0.001$) report such cases. In addition, the student *t*-test statistic is discussed if there was no violation of normality and equality of variances; elsewhere, the Mann–Whitney's U test and Welch's test are reported, respectively. The results suggest statistically significant results of WOM and MTBF (Table 4). However, no statistically significant results were reported considering the impact of outsourced personnel on reducing availability, i.e., MTTR ($p = 0.245$).

Table 4. Independent sample *t*-test statistic of FAP.

Metric	Test	Value	*p*	VS-MPR	SE Diff.	Effect	95%CI$_L$	95%CI$_U$
WOM	Student	3.769	0.001	167.711	188.558	0.715	0.330	1.098
	Welch	3.798	0.001	141.667	187.125	0.718	0.324	1.108
	M-W	1958.5	0.009	8.653		0.272	0.063	0.458
MTTR	Student	1.170	0.245	1.068	0.472	0.222	−0.152	0.595
	Welch	1.174	0.243	1.070	0.471	0.222	−0.151	0.595
	M-W	1654.5	0.496	1.000		0.074	−0.140	0.282
MTBF	Student	−3.585	0.001	95.728	134.985	−0.681	−1.062	−0.296
	Welch	−3.586	0.001	96.211	134.915	−0.681	−1.062	−0.296
	M-W	919	0.001	179.355		−0.403	−0.567	−0.209

Cohen's d reports the effect size of the student *t*-test. For the Mann–Whitney test, the effect size was reported by Rank Biserial Correlation. The Vovk–Sellke Maximum *p*-ratio is based on a two-sided *p*-value as the maximum possible odds in favour of H_1.

Visualised results (Figure 5) suggest a deviation comparing the WOM_Est of indoor and outsourced maintenance for failure diagnosis. Even so, considering the non-parametric results reported via Mann–Whitney ($U = 1985$, $p = 0.009$, RBS = 0.272), a decrease measured by wasted oil monthly (Figure 5a) exists. Next, measuring MTBF (Figure 5b), there is evidence of statistical significance ($t = -3.585$, $p < 0.001$, $d = -0.681$) when comparing indoor and outsourced FAP. This suggests that outsourcing maintenance may positively impact sustainability and improve operational efficiency but will not guarantee an improved time-to-repair metric.

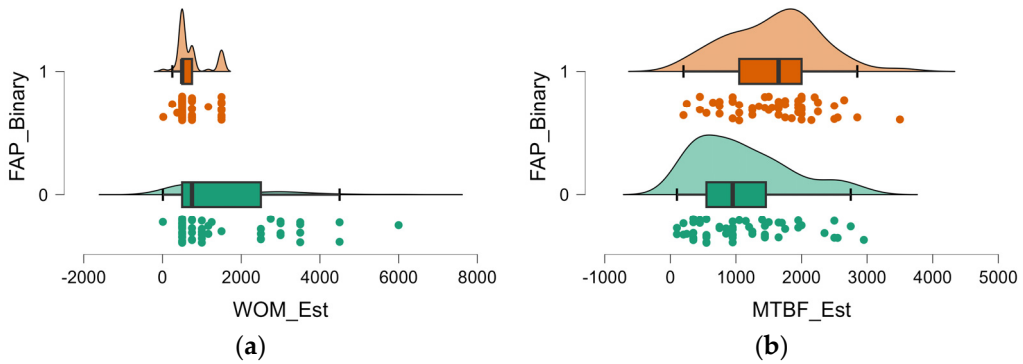


Figure 5. Independent sample statistic of indoor (green) and outdoor (orange) FAP considering (**a**) WOM_Est distributions showing violation of normality ($W = 0.763$, $p < 0.001$) and equality of variance violation using Brown–Forsythe test ($F = 15.442$, $p < 0.001$); and (**b**) MTBF_Est distributions suggesting skewness of normality of indoor FAP ($W = 0.981$, $p = 0.002$) and homogeneity of variances ($F = 0.080$, $p = 0.778$). Note: FAP_Binary (1) = Outdoor; FAP_Binary (0) = Indoor failure analysis.

Considering the analysis of DAT (Table 5), the MTBF reports statistically significant results ($t = -2.519$, $p = 0.025$, $U = 315$, $p = 0.020$) comparing the cases with and without data

analysis tools used in fault detection and diagnosis. When performing the analysis, the lack of a strong effect is presumably due to the presence of an outlier (Figure 6a). Even so, the results suggest slight skewness, although violation in variance homogeneity is not reported. Thus, with caution, we can assume that the results would increase with a larger sample size. The reported effect (RBC = −0.427 with 95%CI [−0.674, −0.062]) and diagnosticity of reported *p*-value = 0.031 using VS-MPR [44,45] suggest a maximum of 3.386 odds in favour of rejecting the null, i.e., in favour of H_1 over H_0. Given the conceptual relation to BF_{10} (Bayes Factor—alternative over null), this would suggest moderate evidence favouring the alternative hypothesis, i.e., using data analysis tools in fault detection and diagnosis improves the time between failures of fluid power systems. However, the key term in VS-MPR is "maximum" odds, which can presumably be ranked between "anecdotal" and "moderate" evidence in the Bayes realm, even though the traditionally reported threshold in frequentist inferential statistics (*p*-value < 0.05) suggests the existence of a statistically significant result.

Table 5. Independent sample *t*-test statistic DAT.

Metric	Test	Value	*p*	VS-MPR	SE Diff.	Effect	95%CI$_L$	95%CI$_U$
WOM	Student	−0.488	0.627	1.000	335.082	−0.155	−0.777	0.468
	Welch	−0.305	0.766	1.000	535.847	−0.116	−0.738	0.511
	M-W.	669.5	0.214	1.116		0.217	−0.139	0.524
MTTR	Student	0.542	0.589	1.000	0.794	0.172	−0.451	0.795
	Welch	0.774	0.450	1.000	0.556	0.200	−0.430	0.823
	M-W.	611.0	0.546	1.000		0.111	−0.245	0.440
MTBF	Student	−2.180	0.031	3.386	233.773	−0.693	−1.320	−0.062
	Welch	−2.519	0.025	3.964	202.317	−0.741	−1.412	−0.048
	M-W.	315.0	0.020	4.624		−0.427	−0.674	−0.095

Cohen's d reports the effect size of the student *t*-test. For the Mann–Whitney test, the effect size was reported by Rank Biserial Correlation. The Vovk–Sellke Maximum *p*-ratio is based on a two-sided *p*-value as the maximum possible odds in favour of H_1.

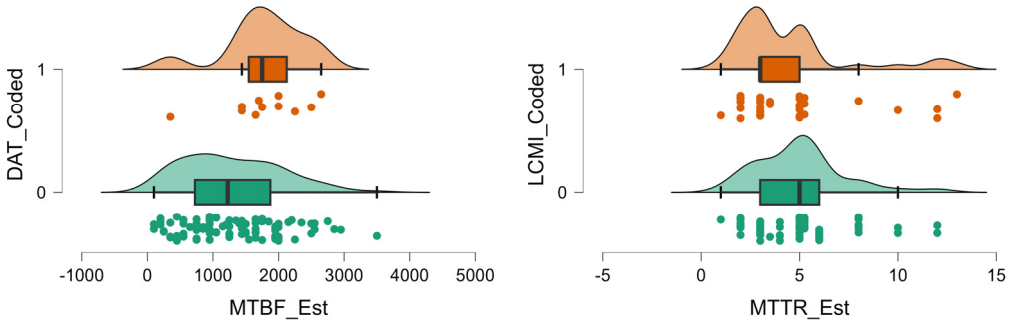


Figure 6. Independent sample test statistic considering (**a**) comparison of MTBF_Est between data analysis tools (1 = Used; 0 = Not used) in fault detection and diagnosis. The Shapiro–Wilk indicates skewness in MTBF_Est_0 ($W = 0.964$, $p < 0.008$) while MTBF_Est_1 ($W = 0.921$, $p < 0.329$) suggests typically distributed values, the Brown–Forsythe test suggests equality of variances. The comparison of (**b**) MTTR_Est considering LCMI (1 = Used; 0 = Not used) indicates the equality of variances ($F = 1.417$, $p = 0.237$) but a violation of normality in both cases ($p < 0.001$).

Regarding the usage of LCMI instruments in terms of performance metrics (Table 6), statistically significant differences are lacking. Namely, although the *t*-test suggests a significant result ($p = 0.037$) in reducing WOM, the Brown–Forsythe test ($F = 3.763$, $p = 0.055$) suggests a potential violation of homogeneity of variances (Table 6), which, in turn, looking at Welch's $t_w = -1.722$, the results indeed fail to reject the null ($p = 0.092$). Finally, looking at the results of MTTR, the non-parametric *U* test = 1663,5 ($p = 0.032$, RBC = 0.251) suggests significant results. However, given the diagnosticity of *p*-value by VS-MPR = 3.312 and

the presence of high skewness (Figure 6b), the results can be inconclusive or rather biased. Relying simply on a non-parametric comparison of ranked scores, the study does not offer strong evidence for rejecting the null.

Table 6. Independent sample *t*-test statistic of LCMI.

Metric	Test	Value	*p*	VS-MPR	SE Diff.	Effect	95%CI$_L$	95%CI$_U$
WOM	Student	−2.107	0.037	2.991	211.453	−0.430	−0.834	−0.025
	Welch	−1.722	0.092	1.676	258.605	−0.385	−0.791	0.026
	M-W.	1154	0.238	1.077		−0.132	−0.350	0.099
MTTR	Student	0.962	0.338	1.003	0.509	0.196	−0.205	0.597
	Welch	0.858	0.395	1.000	0.571	0.185	−0.218	0.586
	M-W.	1663.5	0.032	3.312		0.251	0.024	0.453
MTBF	Student	−1.750	0.083	1.781	151.462	−0.357	−0.760	0.047
	Welch	−1.603	0.115	1.481	165.352	−0.341	−0.745	0.066
	M-W.	1118.5	0.180	1.192		−0.159	−0.374	0.072

Cohen's d reports the effect size of the student *t*-test. For the Mann–Whitney test, the effect size was reported by Rank Biserial Correlation. The Vovk–Sellke Maximum p-ratio is based on a two-sided *p*-value as the maximum possible odds in favour of H_1.

Lastly, from the analysis of laboratorial analysis as a part of LCM practice, the study failed to capture the statistically significant effect (Table 7). Namely, although the *t*-test reports the presence of an effect (p = 0.008, VS-MPR = 9.612), the sample size of only three companies utilising laboratory analysis in their LCM practice does not offer enough evidence to support the findings, primarily since Welch's test reports nonsignificant results (t_w = −0.956, p = 0.440).

Table 7. Independent sample *t*-test statistic of LCML.

Metric	Test	Value	*p*	VSMPR	SE	Effect	95%CI$_L$	95%CI$_U$
WOM	Student	−2.706	0.008	9.612	598.316	−1.584	−2.747	−0.414
	Welch	−0.956	0.440	1.000	1693.95	−0.743	−2.016	0.645
	M-W.	98.50	0.225	1.096		−0.392	−0.793	0.246
MTTR	Student	0.357	0.722	1.000	1.464	0.209	−0.939	1.356
	Welch	0.737	0.522	1.000	0.709	0.267	−0.924	1.411
	M-W.	181.5	0.726	1.000		0.12	−0.496	0.656
MTBF	Student	−0.817	0.416	1.000	438.694	−0.478	−1.626	0.672
	Welch	−0.475	0.681	1.000	754.619	−0.339	−1.491	0.883
	M-W.	131.0	0.579	1.000		−0.191	−0.696	0.439

Cohen's d reports the effect size of the student *t*-test. For the Mann–Whitney test, the effect size was reported by Rank Biserial Correlation. The Vovk–Sellke Maximum *p*-ratio is based on a two-sided *p*-value as the maximum possible odds in favour of H_1.

In conclusion, the analysis of FAP suggests that outsourced personnel for performing fault and failure diagnosis contributes to a reduction in fluid waste and improvement of operational time, but only considering the time between failures. The study lacks evidence that outsourced failure analysis reduces machine downtime, i.e., MTTR. The analysis regarding the impact of data analysis tools on failure analysis suggests moderate evidence considering the positive impact on improving operational time, specifically MTBF. However, the diagnosticity of *p*-values using VS-MPR does not offer strong evidence favouring these findings, although the test statistic suggests significant results. Next, using LCM instruments in fault diagnosis, the analysis suggests a lack of solid evidence suggesting meaningful results. Although the parametric test statistic suggests significant results (p = 0.037), the data are highly skewed, and the normality assumption is violated. Using the non-parametric Mann–Whitney test (p = 0.238), the results suggest a lack of evidence supporting the effect. Lastly, using an expensive laboratorial analysis as a part of LCM practices, the biased sample and limited sample size suggest a lack of an effect measured by maintenance performance metrics.

3.3.2. ANOVA Analysis

The One-way ANOVA analysis is performed on the MDS variable (Table 8). The analysis suggests significant results (p = 0.004, VS-MPR = 18.418). The analysis outcomes show a medium effect size (η^2 = 0.136, ω^2 = 0.103), suggesting that the MDS predictor explains 13.6% of the total variance. The visualisation of categories added further insights into the predictor (Figure 7a). Given the presence of an outlier in the MDS1 category, the evidence suggests conflicting findings and presumably statistically significant results due to the confounding effects of the company size. The ANOVA analysis of MDS categories on the performance of MTBF and MTTR did not report statistically significant results.

Table 8. ANOVA analysis of MDS and WOM.

H-C *	Cases	SS *	df	MS *	*F*	*p*	VSMPR *	η^2	$\eta^2 p$	ω^2
None	MDS_Coded	1.653×10^7	4	4.133×10^6	4.172	0.004	18.418	0.136	0.136	0.103
	Residuals	1.050×10^8	106	990,696.461						
BF *	MDS_Coded	1.653×10^7	4	4.133×10^6	4.665	0.002	26.19	0.136	0.136	0.103
	Residuals	1.050×10^8	63.083	1.665×10^6						
Welch	MDS_Coded	1.653×10^7	4	4.133×10^6	3.945	0.017	5.397	0.136	0.136	0.103
	Residuals	1.050×10^8	19.388	5.416×10^6						

* H-C = Homogeneity Correction; SS = Sum of Squares; MS = Mean Square; BF = Brown–Forsythe; VSMPR = Vovk–Sellke Maximum p-Ratio.

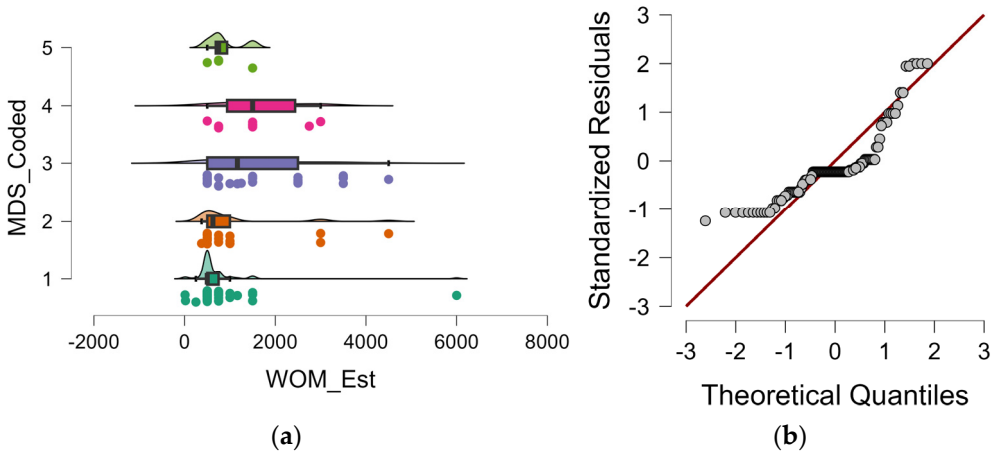


Figure 7. Pairwise comparison of (**a**) MDS categories (1 = Operators, 2 = Technicians, 3 = Engineers, 4 = Managers, and 5 = hydraulic system Specialists) and WOM metric and (**b**) QQ plot of residuals.

Furthermore, the evidence also suggests inequality of variances (F = 4.527, p = 0.002), while the QQ plot suggests deviation of residuals (Figure 7b). Thus, we used the Kruskal–Wallis H test as a non-parametric alternative to the ANOVA. The results report significant findings (H = 23.730, $p < 0.001$). In addition, we perform post hoc pairwise comparisons using Dunn's test (with 1000 bootstrapping) to uncover the difference between categories (Table 9). The results suggest there is a statistically significant difference between MDS1–MDS3 ($p < 0.001$, RBC = −0.429) and MDS1–MDS4 ($p < 0.001$, RBC = −0.733).

Table 9. Post hoc pairwise comparison.

Comparison	z	W_i	W_j	p	p_{bonf}	p_{holm}
1–2	−1.509	42.480	55.083	0.131	1.000	0.788
1–3	−3.843	42.480	69.741	<0.001 ***	0.001 **	0.001 **
1–4	−3.815	42.480	82.650	<0.001 ***	0.001 **	0.001 **
1–5	−1.291	42.480	62.875	0.197	1.000	0.983
2–3	−1.607	55.083	69.741	0.108	1.000	0.756
2–4	−2.300	55.083	82.650	0.021 *	0.215	0.172
2–5	−0.464	55.083	62.875	0.643	1.000	1.000
3–4	−1.158	69.741	82.650	0.247	1.000	0.987
3–5	0.424	69.741	62.875	0.672	1.000	1.000
4–5	1.100	82.650	62.875	0.271	1.000	0.987

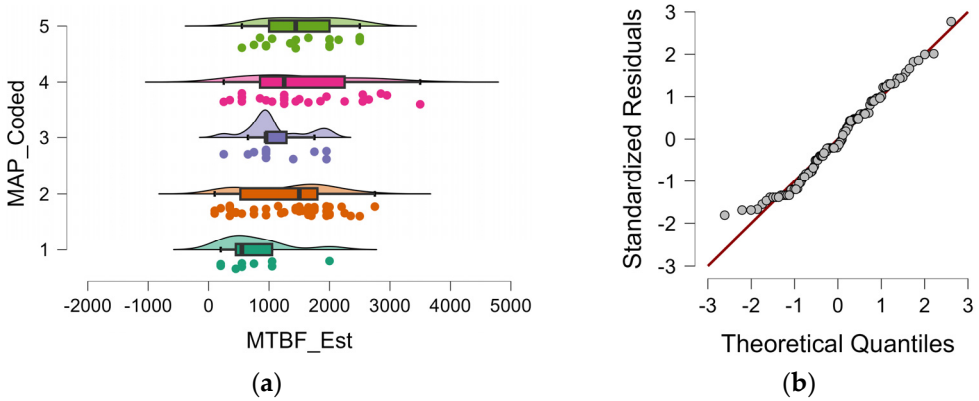
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Considering the MAP effects on performance metrics, the findings are as follows. The influence of MAP on WOM and MTTR performance did not provide significant results. As for the MTBF (Table 10), there is an effect between MAP and MTBF ($p = 0.027$, $\eta^2 = 0.097$), even after homogeneity correction. However, the MAP accounts for about 9.7% of the total variance in MAP. After plotting the results (Figure 8a), the distributions suggest a higher theoretical probability of increasing MTBF with advanced MAP. The normality assumption is not violated (Figure 8b), but Levene's test ($p = 0.022$) suggests heterogeneous variance, which, after correction, the evidence suggests statistically significant results.

Table 10. ANOVA analysis MAP and MTBF.

H-C *	Cases	SS *	df	MS *	F	p	VS-MPR *	η^2 *	$\eta^2 p$ *	ω^2 *
None	MDS_Coded	6.002×10^6	4	1.501×10^6	2.861	0.027	3.778	0.097	0.097	0.063
	Residuals	5.560×10^7	106	524,565.039						
BF *	MDS_Coded	6.002×10^6	4	1.501×10^6	3.347	0.014	6.248	0.097	0.097	0.063
	Residuals	5.560×10^7	81.283	684,075.358						
Welch	MDS_Coded	6.002×10^6	4	1.501×10^6	3.850	0.010	7.697	0.097	0.097	0.063
	Residuals	5.560×10^7	36.127	1.539×10^6						

* H-C = Homogeneity Correction; SS = Sum of Squares; MS = Mean Square; BF = Brown–Forsythe; VSMPR = Vovk–Sellke Maximum p-Ratio. η^2 = the effect size that indicates the proportion of total variance in the dependent variable attributed to an independent variable; $\eta^2 p$ = similar to η^2 but adjusted for other effects; ω^2 = the measure of an effect size adjusted for bias.



(**a**) (**b**)

Figure 8. The distributions with point estimate results of the ANOVA considering (**a**) MAP (1 = None, 2 = Visual Inspection, 3 = Contamination Control, 4 = LCM, 5 = Contamination Control and PHM) categorical predictors (y-axis) and MTBF_Est (x-axis); and (**b**) Quantile–Quantile plot of residuals.

The results suggest the same outcome after performing a post hoc pairwise comparison using parametric Tukey and non-parametric Dunn's test (Table 11) using 1000 bootstrapping. Namely, in both cases, there is a statistically significant difference between MAP1 and

MAP5 categories regarding the performance of MTBF ($t = -2.823$, $p = 0.044$; $z = -2.822$, $p = 0.005$). After using Bonferroni correction on a post hoc Dunn's test, the results report statistically significant results with a large effect size ($p = 0.048$, RBC = −0.721). However, Dunn's test also suggests a significant difference between MAP1 and MAP4 ($p = 0.014$); however, after subjecting the p-value to Bonferroni correction, the results are not ($p = 0.141$). In sum, using advanced maintenance analysis programs, such as contamination control and prognostics and health management [46,47], the MTBF will significantly rise, improving the operational availability of fluid power systems.

Table 11. Post hoc ANOVA test of MAP.

Comparison		Mean Diff.	95%CI$_L$ †	95%CI$_U$ †	SE	Bias	t	d *	p_{tukey} *	p_{bonf} *
1	2	−567.035	−900.745	−17.953	213.502	−5.627	−2.088	−0.764	0.233	0.392
	3	−357.205	−743.86	199.218	223.14	−5.444	−1.102	−0.471	0.805	1.000
	4	−792.467	−1235.963	−200.263	256.796	−3.245	−2.764	−1.074	0.052	0.067
	5	−848.589	−1191.627	−262.425	230.391	−4.888	−2.823	−1.142	0.044	0.057
2	3	219.986	−163.941	528.936	173.093	0.183	0.955	0.293	0.874	1.000
	4	−217.714	−676.854	171.782	213.156	2.383	−1.240	−0.311	0.728	1.000
	5	−268.598	−630.812	61.548	179.665	0.739	−1.379	−0.378	0.643	1.000
3	4	−445.921	−888.478	64.831	232.464	2.199	−1.808	−0.604	0.374	0.734
	5	−487.654	−881.565	−87.631	197.824	0.555	−1.907	−0.672	0.320	0.593
4	5	−46.923	−488.685	390.18	226.334	−1.644	−0.223	−0.068	0.999	1.000

† Bias corrected accelerated (95% lower and upper control limits). * d = Cohen's d statistic; p-value adjusted using Tukey and Bonferroni correction.

Lastly, the analysis of CMS and MTBF_Est (Table 12) and the ANOVA analysis, suggests the presence of significant findings ($F = 4.369$, $\eta^2 = 0.109$). At the same time, CMS did not have an apparent effect on WOM and MTTR performance metrics. After performing the normality analysis, the distributions of CMS categories (Figure 9a) and the QQ plot (Figure 9b) suggest that normality was not violated. Next, the analysis of Levene's test ($p = 0.897$) suggests homogeneity of variances. Consequently, a Tukey post hoc pairwise comparison is performed (Table 13).

Table 12. The ANOVA analysis of CMS and MTBF_Est.

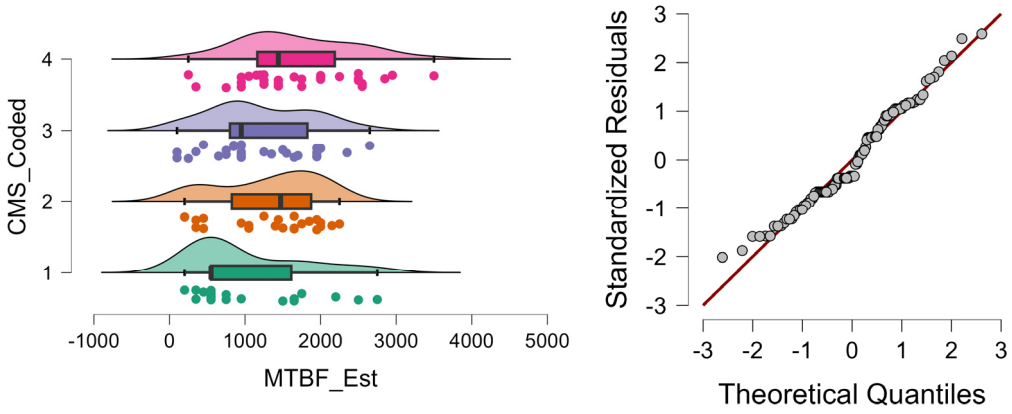
H-C *	Cases	SS *	df	MS *	F	p	VS-MPR *	η^2	$\eta^2 p$	ω^2
None	MDS_Coded	6.724×10^6	3	2.241×10^6	4.369	0.006	11.878	0.109	0.109	0.083
	Residuals	5.488×10^7	107	512,919.89						
BF *	MDS_Coded	6.724×10^6	3	2.241×10^6	4.383	0.006	11.818	0.109	0.109	0.083
	Residuals	5.488×10^7	99.414	552,058.505						
Welch	MDS_Coded	6.724×10^6	3	2.241×10^6	3.885	0.014	6.303	0.109	0.109	0.083
	Residuals	5.488×10^7	56.059	979,014.002						

* H-C = Homogeneity Correction; SS = Sum of Squares; MS = Mean Square; BF = Brown–Forsythe; VSMPR = Vovk–Sellke Maximum p-Ratio.

The post hoc analysis (Table 13) suggests a statistically significant relationship only between CMS1 and CMS4 with a substantial effect size ($p_{tukey} = 0.005$, $p_{bonferroni} = 0.006$, $d = -0.928$). This, in turn, suggests that, although more sophisticated condition monitoring sensors affect the MTBF, statistically significant results are reported only for maintenance practices that utilise and rely on a complete set of sensors (e.g., pressure, flow, temperature, contamination, and vibroacoustic).

In sum, the influence of maintenance department staff—operators, technicians, engineers, and hydraulic system specialists offer conflicting findings in improving sustainability by the amount of waste oil monthly, presumably due to confounding effects. The performance metrics used to measure overall equipment efficiency through availability, specifically MTTR and MTBF, do not offer statistically significant results. The analysis of maintenance analysis programs, for instance, using advanced maintenance programs

(e.g., contamination control), significantly improves the MTBF of fluid power systems. In contrast, the impact on other performance metrics was not documented. Lastly, using instruments for condition monitoring of fluid power systems proposes similar findings. Specifically, the post hoc pairwise comparison suggests the difference between the CMS1 and CMS4 ($t_{bonferroni} = -3.391$, $p_{bonferroni} = 0.006$), ultimately showing that contamination control instruments affect the fluid power system effectiveness, particularly in increasing the MTBF metric.



(**a**) (**b**)

Figure 9. The distributions with point estimates of (**a**) CMS (1 = None, 2 = Pressure, 3 = Pressure, Flow and Temperature, 4 = Pressure, Flow, Temperature and Contamination Control) (*y*-axis) and MTBF_Est (*x*-axis); and (**b**) Quantile–Quantile plot of residuals.

Table 13. Post hoc ANOVA test of CMS.

Comparison		Mean Diff.	95%CI$_L$ †	95%CI$_U$ †	SE	Bias	*t*	*d* *	p_{tukey} *	p_{bonf} *
1	2	−300.638	−695.270	116.112	205.510	3.273	−1.434	−0.423	0.481	0.928
	3	−199.358	−564.709	216.667	200.520	3.055	−1.013	−0.282	0.742	1.000
	4	−649.846	−1095.630	−280.948	205.606	6.667	−3.391	−0.928	0.005	0.006
2	3	105.039	−281.666	459.568	184.677	−0.218	0.518	0.141	0.955	1.000
	4	−351.316	−770.825	−2.045	191.945	3.394	−1.894	−0.505	0.237	0.366
3	4	−455.082	−855.195	−108.529	183.477	3.612	−2.600	−0.646	0.051	0.064

† Bias corrected accelerated (95% lower and upper control limits). * *d* = Cohen's d statistic; *p*-value adjusted using Tukey and Bonferroni correction.

3.3.3. Correlation Analysis

The parametric Pearson's and non-parametric Spearman's correlations are performed as the last step in the inferential statistical analysis of features. In the preliminary study, we first perform multivariate and bivariate normality testing. The results suggest a violation of multivariate normality ($p < 0.001$), while in bivariate normality, only the relationship between time to oil refilling (TTOR) and MTBF did not violate the normality assumption ($p = 0.149$). Therefore, based on Spearman's correlation, the TTOR and FRT are eliminated as they do not provide significant results ($p < 0.05$). In the following, we provide Spearman's heatmap (Figure 10) for a more straightforward interpretation and discussion of the results.

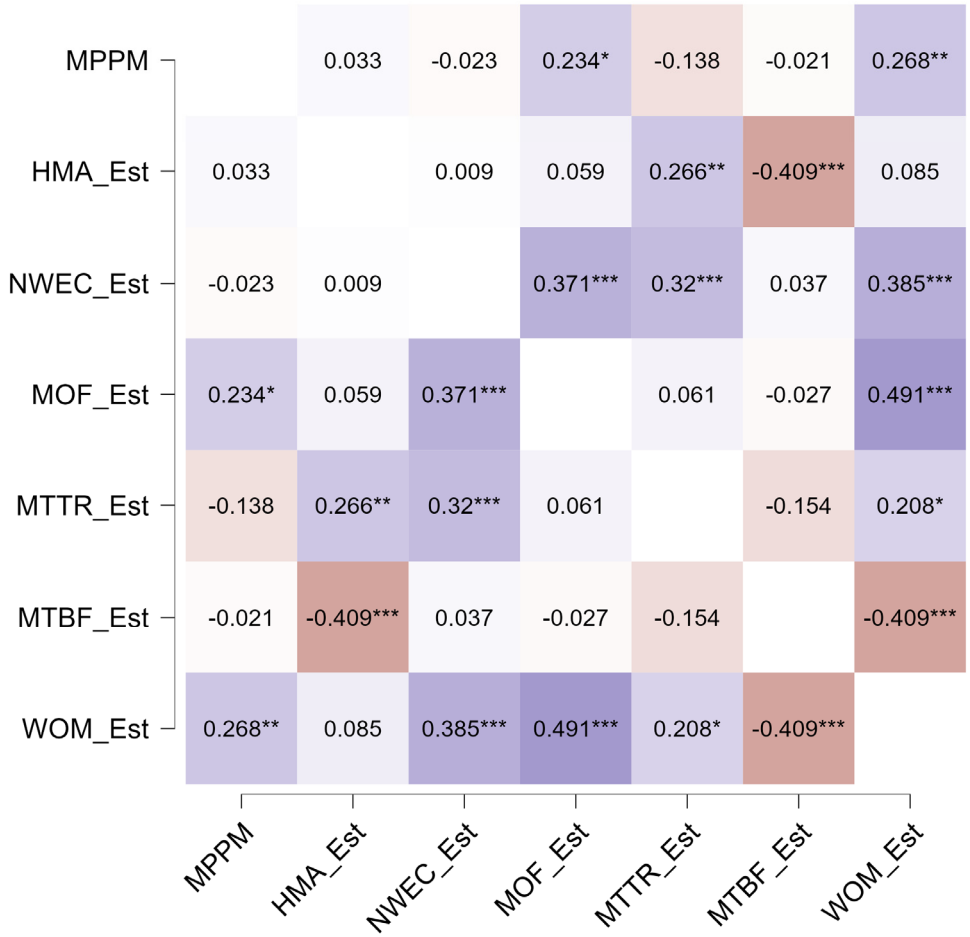


Figure 10. Spearman's correlation heatmap (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; blue indicating positive correlation and red indicating negative correlation).

The MPPM (Maintenance Personnel Per Machine) metric, which describes the number of personnel responsible for the maintenance of assets, suggests a positive correlation with the WOM metric ($\rho = 0.234$, $p = 0.013$, $z = 0.225$). As previously reported, this may be attributed to the confounding effect of company size and type. Next, the assets' age suggests a positive correlation with MTTR ($\rho = 0.266$, $p = 0.01$) and a negative correlation with MTBF ($\rho = -0.409$, $p \leq 0.001$). This suggests a significant increase in machine downtime, while in contrast, reduced time between failures, ultimately confirming that as an asset gets older, the availability decreases, ultimately affecting equipment effectiveness. The NWEC predictor, which is derived as a fluid power system capacity expressed via nominal working power (in [kW] power unit), results in a positive correlation with MTTR ($\rho = 0.320$, $p \leq 0.001$) and WOM ($\rho = 0.385$, $p \leq 0.001$). This, in turn, shows that fluid power systems with higher capacity tend to consume repair time and produce excessive fluid waste, which is expected in practice. Interestingly, there was no effect between NWEC and MTBF, which was surprising given that the literature often reports an increased failure rate as the capacity of a power system is higher [48]. Similarly, as the oil capacity or machine oil filling tends to be related to the power capacity of fluid power systems, the positive correlation with WOM is reported ($\rho = 0.491$, $p \leq 0.001$), which was expected. In sum, the obtained findings do not deviate from practical understanding considering proposed continuous predictors; they further emphasise the problem of dealing with EoL (End of Life) issues in industrial manufacturing and service industrial machinery [49]. However, novel conclusions are obtained regarding features not commonly analysed in practice and will be discussed in the following.

4. Conclusions

4.1. Concluding Remarks

The analysis relies on questionnaire-based survey data to identify the main maintenance factors affecting the performance of industrial assets, which rely on model-agnostic interpretation. Given that three target (dependent) variables are collected via survey, we used PCA (Principal Component Analysis) to reduce the dimensionality of response variables. The first component is used for ML regression. After obtaining features by the mean dropout loss, the same is used for hypothesis testing to estimate individual target variables' actual contribution.

The findings suggest that outsourced failure analysis offers the highest impact on improving availability, specifically increasing Mean Time Between Failures (MTBF). Next, the data analysis tool utilised alongside advanced condition monitoring sensors suggests a positive impact on improving the MTBF of fluid power systems. The analysis of maintenance analysis program predictors also supports previous claims, suggesting that using advanced maintenance analysis programs, such as contamination control and prognostics and health management, improves operational effectiveness expressed via MTBF. Finally, the fluid power system's age (i.e., HMA) significantly impacts the negative correlation effect on MTBF, which is commonly reported.

Considering MTTR performance, the HMA estimate positively correlates with MTTR, suggesting that age significantly impacts the time needed to return the system to its operational state. This has been supported by the capacity of fluid power systems expressed in [kW] unit size, suggesting that larger machines are associated with longer repair times—MTTR. Using lubricant condition monitoring instruments reports a potential impact on MTTR; however, the results may be uncertain due to the limited size and violation of the test's assumption.

Lastly, the impact on sustainability performance measured by the wasted hydraulic oil per month suggests that outsourced failure analysis improves sustainability performance expressed in monthly waste oil. The findings also highlight the cases where maintenance departments consisting of managers and hydraulic specialists will positively impact sustainability, i.e., reducing monthly oil waste. However, it should be noted that with the increase in company size, maintenance personnel and machine size (expressed in kW unit power), there is a higher risk of increasing sustainability impact in terms of waste oils. Although using lubricant condition monitoring instruments and performing oil laboratory analysis suggests the existence of a statistically significant impact, the impact of these variables remains inconclusive due to the violation of several assumptions and limited sample size.

4.2. Limitations of the Study

Several limitations were identified in the study. Firstly, the resulting dimensionality reduction from using Principal Component Analysis (PCA) suggests that the first component accounts for about 49% of the overall variance. The commonly used combination with PCA relies on >70% of the variance explained. However, this certainly does not downplay our findings given that we used dimensionality reduction, i.e., PCA, only on the first component; combining the first and second components accounts for 80% of the variance. In addition, we then performed a hypothesis test by assessing each variable on all three target variables to ensure the reliability and validity of outcomes.

Secondly, some predictor and target variables synthesised from the survey suggest 1–4% missing data, imputed using average values or mean imputation. This, in turn, slightly reduced the data variability, leading to underestimation of standard errors and overestimation of significance in predictor coefficients. That is why, in our hypothesis testing and significance analysis, we performed bootstrapping and sensitivity analysis to ensure the actual impact of predictors on individual target variables. This helped us increase the robustness and generalisability of the findings.

Finally, the findings obtained may suffer from heterogeneity in applications. Since fluid power systems are employed in different working environments and regimes, a

hydraulic subsystem (fluid power system) in an extrusion production line and a mining excavator significantly differ in their applications. Therefore, the findings may benefit from isolating the target organisations to a specific application that will offer a more homogeneous study sample.

4.3. Implications and Future Research

From a practical standpoint, the results may impact companies' policymaking in considering the integration of advanced digital technologies and digital competencies, as can be observed by the benefits and, at the same time, the lack of advanced data analytic tools and maintenance programs focused on prognostics and health management. This is not only for operational and economic efficiency, as observed by improving availability through the MTBF but also for sustainability reasoning. Furthermore, the significant impact of machine age on fluid power system performance, observed by the impact on availability metrics, offers a range of possibilities for incorporating advanced maintenance practices, such as Energy-Based Maintenance and Sustainable Predictive Maintenance dedicated to End-of-Life issues [50]. Lastly, and most importantly, it seems that maintenance is at the intersection of becoming a MaaS (Maintenance as a Service) function, given that the study reports that outsourcing maintenance failure analysis significantly contributes to the overall process improvement and sustainability. Namely, it has been observed that outsourcing failure analysis may be beneficial in terms of time between failures and waste oil. In conclusion, the companies may consider altering their maintenance practices and incorporating more advanced digital solutions to improve these performance metrics, which is where our future research will reside. We will assess the impact on maintenance performance metrics by investigating such cases and contrasting the findings to see the results.

Author Contributions: Conceptualisation, M.O.; Methodology, M.O.; Validation, M.O., I.B., N.B. and V.V.; Formal analysis, M.O. and N.B.; Investigation, M.O., I.B., N.B. and V.V.; Resources, M.O. and I.B.; Data curation, M.O., N.B. and V.V.; Writing—original draft preparation, M.O. and V.V.; Writing—review and editing, M.O., I.B., N.B. and V.V.; Visualisation, M.O.; Supervision, I.B. and N.B.; Project administration, M.O., I.B., V.V. and N.B.; Funding acquisition, M.O., I.B., N.B. and V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-65/2024-03/200156), and the Faculty of Technical Sciences, University of Novi Sad, through the project "Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad" (No. 01-3394/1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nurprihatin, F.; Angely, M.; Tannady, H. Total Productive Maintenance Policy to Increase Effectiveness and Maintenance Performance Using Overall Equipment Effectiveness. *J. Appl. Res. Ind. Eng.* **2019**, *6*, 184–199. [CrossRef]
2. Pires, S.d.P.; Sénéchal, O.; Deschamps, F.; Loures, E.R.; Perroni, M.G. Industrial Maintenance for Sustainable Performance: A Systematic Literature Review. In Proceedings of the 23rd International Conference on Production Research, Manila, Philippines, 2–5 August 2015.
3. Roda, C.; Voisin, I.; Miranda, A.; Macchi, S.; Iung, M. *Sustainable Maintenance Performances and EN 15341: 2019: An Integration Proposal*; Springer: Berlin/Heidelberg, Germany, 2021.
4. Werbińska-Wojciechowska, S.; Winiarska, K. Maintenance Performance in the Age of Industry 4.0: A Bibliometric Performance Analysis and a Systematic Literature Review. *Sensors* **2023**, *23*, 1409. [CrossRef] [PubMed]
5. Orošnjak, M.; Brkljač, N.; Šević, D.; Čavić, M.; Oros, D.; Penčić, M. From Predictive to Energy-Based Maintenance Paradigm: Achieving Cleaner Production through Functional-Productiveness. *J. Clean. Prod.* **2023**, *408*, 137177. [CrossRef]

6. Jasiulewicz-Kaczmarek, M. The Role of Ergonomics in Implementation of the Social Aspect of Sustainability, Illustrated with the Example of Maintenance. In *Occupational Safety and Hygiene*; CRC Press: Boca Raton, FL, USA, 2013; pp. 61–66.
7. Jasiulewicz-Kaczmarek, M.; Antosz, K. Industry 4.0 Technologies for Maintenance Management—An Overview. In Proceedings of the International Conference Innovation in Engineering, Minho, Portugal, 28–30 June 2022; Lecture Notes in Mechanical Engineering. Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2022; pp. 68–79.
8. Karuppiah, K.; Sankaranarayanan, B.; Ali, S.M. On Sustainable Predictive Maintenance: Exploration of Key Barriers Using an Integrated Approach. *Sustain. Prod. Consum.* **2021**, *27*, 1537–1553. [CrossRef]
9. Turner, C.; Okorie, O.; Oyekan, J. XAI Sustainable Human in the Loop Maintenance. In Proceedings of the IFAC-PapersOnLine, Casablanca, Morocco, 29 June–1 July 2022; Elsevier B.V.: Amsterdam, The Netherlands, 2022; Volume 55, pp. 67–72.
10. Franciosi, C.; Di Pasquale, V.; Iannone, R.; Miranda, S. A Taxonomy of Performance Shaping Factors for Human Reliability Analysis in Industrial Maintenance. *J. Ind. Eng. Manag.* **2019**, *12*, 115–132. [CrossRef]
11. Orošnjak, M.; Šević, D. Benchmarking Maintenance Practices for Allocating Features Affecting Hydraulic System Maintenance: A West-Balkan Perspective. *Mathematics* **2023**, *11*, 3816. [CrossRef]
12. Sari, E.; Shaharoun, A.M.; Ma'aram, A.; Yazid, A.M. Sustainable Maintenance Performance Measures: A Pilot Survey in Malaysian Automotive Companies. *Procedia CIRP* **2015**, *26*, 443–448. [CrossRef]
13. Orosnjak, M. Maintenance Practice Performance Assessment of Hydraulic Machinery: West Balkan Meta-Statistics and Energy-Based Maintenance Paradigm. In Proceedings of the 2021 5th International Conference on System Reliability and Safety (ICSRS), Palermo, Italy, 24–26 November 2021; IEEE: Piscataway, NJ, USA; pp. 108–114.
14. Judijanto, L.; Uhai, S.; Suri, I. The Influence of Business Analytics and Big Data on Predictive Maintenance and Asset Management. *Eastasouth J. Inf. Syst. Comput. Sci.* **2024**, *1*, 123–135. [CrossRef]
15. Díaz-Reza, J.; García-Alcaraz, J.; Avelar-Sosa, L.; Mendoza-Fong, J.; Sáenz Diez-Muro, J.; Blanco-Fernández, J. The Role of Managerial Commitment and TPM Implementation Strategies in Productivity Benefits. *Appl. Sci.* **2018**, *8*, 1153. [CrossRef]
16. Orošnjak, M.; Delić, M.; Ramos, S. Influence of Maintenance Practice on MTBF of Industrial and Mobile Hydraulic Failures: A West Balkan Study. In *Machine and Industrial Design in Mechanical Engineering*; Springer: Cham, Switzerland, 2022; pp. 617–625.
17. Pineda, J.P. Case Study: Characterizing Life Expectancy Drivers across Countries Using Model-Agnostic Interpretation Methods for Black-Box Models. Available online: https://rpubs.com/JoPaPi/1066511 (accessed on 28 May 2024).
18. Greenacre, M.; Groenen, P.J.F.; Hastie, T.; D'Enza, A.I.; Markos, A.; Tuzhilina, E. Publisher Correction: Principal Component Analysis. *Nat. Rev. Methods Primers* **2023**, *3*, 22. [CrossRef]
19. Peres-Neto, P.R.; Jackson, D.A.; Somers, K.M. Giving Meaningful Interpretation to Ordination Axes: Assessing Loading Significance in Principal Component Analysis. *Ecology* **2003**, *84*, 2347–2363. [CrossRef]
20. Bro, R.; Smilde, A.K. Principal Component Analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [CrossRef]
21. Abdi, H.; Williams, L.J. Principal Component Analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
22. Chandrvanshi, S.; Sharma, S.; Singh, M.P.; Singh, R. Bearing Fault Diagnosis Using Machine Learning Models. In *Micro-Electronics and Telecommunication Engineering*; Springer: Cham, Switzerland, 2024; pp. 219–233.
23. Cui, B.; Weng, Y.; Zhang, N. A Feature Extraction and Machine Learning Framework for Bearing Fault Diagnosis. *Renew. Energy* **2022**, *191*, 987–997. [CrossRef]
24. Cartocci, N.; Napolitano, M.R.; Crocetti, F.; Costante, G.; Valigi, P.; Fravolini, M.L. Data-Driven Fault Diagnosis Techniques: Non-Linear Directional Residual vs. Machine-Learning-Based Methods. *Sensors* **2022**, *22*, 2635. [CrossRef]
25. Olive, D.J. Multiple Linear Regression. In *Linear Regression*; Springer International Publishing: Cham, Switzerland, 2017; pp. 17–83.
26. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. Linear Regression. In *An Introduction to Statistical Learning*; Springer: Cham, Switzerland, 2023; pp. 69–134.
27. Abu Amra, I.A.; Maghari, A.Y.A. Students Performance Prediction Using KNN and Naïve Bayesian. In Proceedings of the 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan, 17–18 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 909–913.
28. Lu, J.; Qian, W.; Li, S.; Cui, R. Enhanced K-Nearest Neighbor for Intelligent Fault Diagnosis of Rotating Machinery. *Appl. Sci.* **2021**, *11*, 919. [CrossRef]
29. Lu, Q.; Shen, X.; Wang, X.; Li, M.; Li, J.; Zhang, M. Fault Diagnosis of Rolling Bearing Based on Improved VMD and KNN. *Math. Probl. Eng.* **2021**, *2021*, 2530315. [CrossRef]
30. Gyamerah, S.A.; Ngare, P.; Ikpe, D. Probabilistic Forecasting of Crop Yields via Quantile Random Forest and Epanechnikov Kernel Function. *Agric. Meteorol.* **2020**, *280*, 107808. [CrossRef]
31. Sun, Y.; Peng, G. Developing Area Real Estate Valuation Based on Linear Regression and KNN Algorithm. In Proceedings of the 2022 6th Annual International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, 14–18 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 38–42.
32. Kherif, O.; Benmahamed, Y.; Teguar, M.; Boubakeur, A.; Ghoneim, S.S.M. Accuracy Improvement of Power Transformer Faults Diagnostic Using KNN Classifier With Decision Tree Principle. *IEEE Access* **2021**, *9*, 81693–81701. [CrossRef]
33. Xueli, W.; Zhiyong, J.; Dahai, Y. An Improved KNN Algorithm Based on Kernel Methods and Attribute Reduction. In Proceedings of the 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), Qinhuangdao, China, 18–20 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 567–570.
34. Bailey, T.M. Convergence of Rprop and Variants. *Neurocomputing* **2015**, *159*, 90–95. [CrossRef]

35. Anastasiadis, A.D.; Magoulas, G.D.; Vrahatis, M.N. New Globally Convergent Training Scheme Based on the Resilient Propagation Algorithm. *Neurocomputing* **2005**, *64*, 253–270. [CrossRef]
36. Hinz, T.; Navarro-Guerrero, N.; Magg, S.; Wermter, S. Speeding up the Hyperparameter Optimization of Deep Convolutional Neural Networks. *Int. J. Comput. Intell. Appl.* **2018**, *17*, 1850008. [CrossRef]
37. Toma, R.N.; Prosvirin, A.E.; Kim, J.-M. Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers. *Sensors* **2020**, *20*, 1884. [CrossRef]
38. Shakiba, F.M.; Azizi, S.M.; Zhou, M.; Abusorrah, A. Application of Machine Learning Methods in Fault Detection and Classification of Power Transmission Lines: A Survey. *Artif. Intell. Rev.* **2023**, *56*, 5799–5836. [CrossRef]
39. Kumar, P.; Hati, A.S. Review on Machine Learning Algorithm Based Fault Detection in Induction Motors. *Arch. Comput. Methods Eng.* **2021**, *28*, 1929–1940. [CrossRef]
40. Donnelly, N.; Cunningham, A.; Salas, S.M.; Bracher-Smith, M.; Chawner, S.; Stochl, J.; Ford, T.; Raymond, F.L.; Escott-Price, V.; van den Bree, M.B.M. Identifying the Neurodevelopmental and Psychiatric Signatures of Genomic Disorders Associated with Intellectual Disability: A Machine Learning Approach. *Mol. Autism* **2023**, *14*, 19. [CrossRef] [PubMed]
41. Éliás, S.; Wrzodek, C.; Deane, C.M.; Tissot, A.C.; Klostermann, S.; Ros, F. Prediction of Polyspecificity from Antibody Sequence Data by Machine Learning. *Front. Bioinform.* **2024**, *3*, 1286883. [CrossRef] [PubMed]
42. Farahani, H.; Blagojević, M.; Azadfallah, P.; Watson, P.; Esrafilian, F.; Saljoughi, S. Feature Selection in AP. In *An Introduction to Artificial Psychology*; Springer International Publishing: Cham, Switzerland, 2023; pp. 187–205.
43. Van Horenbeek, A.; Pintelon, L.; Muchiri, P. Maintenance Optimization Models and Criteria. *Int. J. Syst. Assur. Eng. Manag.* **2010**, *1*, 189–200. [CrossRef]
44. Sellke, T.; Bayarri, J.M.; Berger, J.O. Calibration of p Values for Testing Precise Null Hypotheses. *Am. Stat.* **2001**, *55*, 62–71. [CrossRef]
45. Vovk, V.G. A Logic of Probability, with Application to the Foundations of Statistics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1993**, *55*, 317–351. [CrossRef]
46. Orošnjak, M.; Jocanović, M.; Karanović, V. Applying Contamination Control for Improved Prognostics and Health Management of Hydraulic Systems. In *Advances in Asset Management and Condition Monitoring. Smart Innovation, Systems and Technologies*; Springer: Cham, Switzerland, 2020; pp. 583–596.
47. Wakiru, J.M.; Pintelon, L.; Karanović, V.V.; Jocanović, M.T.; Orošnjak, M.D. Analysis of Lubrication Oil towards Maintenance Grouping for Multiple Equipment Using Fuzzy Cluster Analysis. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *393*, 012011. [CrossRef]
48. Postnikov, S.E.; Trofimov, A.A.; Baikov, S.V. Architecture Options Estimate for the Near-Medium-Haul Aircraft Control System by the Reliability, Mass and Power Consumption Criteria. *Aerosp. Syst.* **2019**, *2*, 33–40. [CrossRef]
49. Santos, A.C.d.J.; Cavalcante, C.A.V.; Wu, S. Maintenance Policies and Models: A Bibliometric and Literature Review of Strategies for Reuse and Remanufacturing. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 108983. [CrossRef]
50. Jasiulewicz-Kaczmarek, M.; Legutko, S.; Kluk, P. Maintenance 4.0 Technologies—New Opportunities for Sustainability Driven Maintenance. *Manag. Prod. Eng. Rev.* **2020**, *11*, 74–87. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.