

Article

Impacts of Feature Selection on Predicting Machine Failures by Machine Learning Algorithms

Francisco Elânio Bezerra ¹, Geraldo Cardoso de Oliveira Neto ², Gabriel Magalhães Cervi ³,
Rafaella Francesconi Mazetto ³, Aline Mariane de Faria ³, Marcos Vido ⁴, Gustavo Araujo Lima ⁵,
Sidnei Alves de Araújo ⁵, Mauro Sampaio ⁶ and Marlene Amorim ^{7,*}

- ¹ Department of Energy Engineering and Electrical Automation, Polytechnic School, University of São Paulo (USP), 158 Prof. Luciano Gualberto Avenue, São Paulo 05508-010, Brazil; elanio@usp.br
- ² Industrial Engineering Post Graduation Program, Federal University of ABC, Alameda da Universidade, s/nº Bairro Anchieta, São Bernardo do Campo, São Paulo 09606-045, Brazil; geraldo.prod@gmail.com
- ³ Business Administration Post-Graduation Program, FEI University, Tamandaré Street 688, 5 Floor, São Paulo 01525-000, Brazil; engineer cervi@gmail.com (G.M.C.); rmazetto@fei.edu.br (R.F.M.); alinefaria@fei.edu.br (A.M.d.F.)
- ⁴ Industrial Engineering Post-Graduation Program, Nove de Julho University (UNINOVE), Vergueiro Street 235/249, São Paulo 01504-001, Brazil; marcosvido8@gmail.com
- ⁵ Informatics and Knowledge Management Post-Graduation Program, Nove de Julho University (UNINOVE), Vergueiro Street 235/249, São Paulo 01504-001, Brazil; gustavo.araujo.lima94@gmail.com (G.A.L.); saraujo@uni9.pro.br (S.A.d.A.)
- ⁶ Industrial Engineering Post-Graduation Program, FEI University, Avenue Humberto de Alencar Castelo Branco 3972-B, São Bernardo do Campo, Assunção 09850-901, Brazil; msampaio@fei.edu.br
- ⁷ GOVCOPP-DEGEIT, University of Aveiro, 3810-193 Aveiro, Portugal
- * Correspondence: mamorim@ua.pt; Tel.: +351-234-370-200



Citation: Bezerra, F.E.; Oliveira Neto, G.C.d.; Cervi, G.M.; Francesconi Mazetto, R.; Faria, A.M.d.; Vido, M.; Lima, G.A.; Araújo, S.A.d.; Sampaio, M.; Amorim, M. Impacts of Feature Selection on Predicting Machine Failures by Machine Learning Algorithms. *Appl. Sci.* **2024**, *14*, 3337. <https://doi.org/10.3390/app14083337>

Academic Editor: Tibor Krenicky

Received: 24 February 2024

Revised: 5 April 2024

Accepted: 11 April 2024

Published: 15 April 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/>).

Abstract: In the context of Industry 4.0, managing large amounts of data is essential to ensure informed decision-making in intelligent production environments. It enables, for example, predictive maintenance, which is essential for anticipating and identifying causes of failures in machines and equipment, optimizing processes, and promoting proactive management of human, financial, and material resources. However, generating accurate information for decision-making requires adopting suitable data preprocessing and analysis techniques. This study explores the identification of machine failures based on synthetic industrial data. Initially, we applied the feature selection techniques Principal Component Analysis (PCA), Minimum Redundancy Maximum Relevance (mRMR), Neighborhood Component Analysis (NCA), and Denoising Autoencoder (DAE) to the collected data and compared their results. In the sequence, a comparison among three widely known machine learning classifiers, namely Random Forest (RF), Support Vector Machine (SVM), and Multilayer Perceptron neural network (MLP), was conducted, with and without considering feature selection. The results showed that PCA and RF were superior to the other techniques, allowing the classification of failures with rates of 0.98, 0.97, and 0.98 for the accuracy, precision, and recall metrics, respectively. Thus, this work contributes by solving an industrial problem and detailing techniques to identify the most relevant variables and machine learning algorithms for predicting machine failures that negatively impact production planning. The findings provided by this study can assist industries in giving preference to employing sensors and collecting data that can contribute more effectively to machine failure predictions.

Keywords: machine learning; machine failure; feature selection; predictive maintenance; sensor selection

1. Introduction

The increasing use of sensors in industries, driven by the rise of intelligent sensor systems in Industry 4.0, represents a significant transformation in production and manufacturing processes. These sensors connect devices and systems, enabling machine-to-machine communication to monitor systems and equipment in industrial facilities. Sensors improve product quality, reduce production costs, and increase operational efficiency by processing data locally and making fast, accurate decisions. Furthermore, the evolution of sensor technologies and their integration with other innovations, such as big data, artificial intelligence, and cloud computing, drives Industry 4.0 towards smarter and more automated production. This paradigm shift in the industry offers promising commercial opportunities as sensors become essential in driving innovation and market competitiveness [1].

In this scenario, machines and managers face daily challenges related to massive data entry and customization in the manufacturing process. Thus, predictive maintenance is a fundamental approach to anticipating failures through advanced analysis, optimizing process efficiency, and promoting proactive resource management, contributing to operational excellence in manufacturing operations [2–6].

The use of advanced techniques to develop prediction models has become crucial to taking advantage of the growing volume of data collected in industries. Machine learning techniques shine in this scenario, using algorithms to examine data in real time and predict output values [7]. It enables meticulous analysis and empowers strategic decision-making based on extensive data sets, addressing the challenges of variety, speed, and volume of information collected in industries [2].

Considering this, numerous studies have used artificial intelligence techniques to extract valuable insights, helping to make decisions related to industry machine failures. In [3], the authors have synthesized the most recent works on predictive maintenance, highlighting that 32 studies used actual data to develop predictive models related to maintenance, while two opted for synthetic data. The analysis reveals that 33% of researchers chose Random Forest (RF), 27% opted for Artificial Neural Networks (ANNs), 25% decided on Support Vector Machines (SVMs), and 13% used K-Means.

Feature selection is a fundamental tool in data analysis, machine learning, and data mining, as it can identify the most relevant features, allowing a deeper understanding of the problem under study and helping data professionals and computer scientists understand which aspects of the data have the most significant impact on the model's predictions. Furthermore, by highlighting the most important features, it is possible to simplify the model, reducing its complexity and making it more computationally efficient [2–5,8]. This step involves choosing the most relevant or informative dataset characteristics to be used in model construction, improving model performance, and reducing dimensionality, interpretability, and computational savings, among other benefits [9–12].

Choosing adequate techniques for feature selection and developing prediction models for machine failure classification are critical points in providing accurate information that assists decision-making processes. Furthermore, these techniques could be applied with the aim of identifying the most useful sensors and providing relevant information for analysis. There are still many other questions underexplored in the literature on automatic identification of industrial machine failures. In this context, this study is guided by the following research questions:

1. Which feature selection techniques are most used for the machine failure prediction problem?
2. Which feature selection techniques will provide more accurate models for predicting machine failure?
3. Which machine learning techniques are most adequate for the machine failure prediction problem?
4. Can feature selection and machine learning help decision-makers choose the most appropriate sensors for acquiring machine data?

Faced with these questions, the Scopus database was used to research these topics, using the keywords “machine learning”, “predictive maintenance”, “feature selection”,

“machine failure classification”, “Random Forest” or “Support Vector Machine” or “Artificial Neural Network”. As a result, 15 scientific articles were found, of which 7 are complete articles, 6 are conference papers, and 2 are conference reviews. Only seven articles have returned when selecting only the “article” document type.

Therefore, this work presents an approach based on machine learning and feature selection techniques to improve the accuracy of the classification of failure machines, aiding predictive maintenance. This work will offer the following contributions:

- Employ and compare feature selection techniques to improve the accuracy of failure classification by utilizing the most important features.
- Present the most relevant feature selection technique for the failure classification problem.
- Test different configurations for the three most used machine learning techniques in the literature, according to [4], and point out which technique improves the accuracy of failure classification using selected features.
- Explore how attribute selection can assist in choosing the most appropriate sensors for acquiring data about machines.

2. Literature Review

2.1. Feature Selection

Feature selection is the process of identifying a valuable subset of features from the original dataset to be used in building a model. Feature selection techniques aim to reduce the dimensionality of the feature space, eliminate redundant data, enhance data quality, improve model accuracy, and gain knowledge about the process that led to the data [13–15].

Three approaches for feature selection exist: supervised, unsupervised, and semi-supervised. Supervised techniques select relevant features based on labeled datasets using filter, wrapper, or embedded models. Filter models evaluate features independently of the classification model; wrapper models interact with the classification model during feature selection; and embedded models incorporate feature selection into classifier construction. Unsupervised techniques, such as Principal Component Analysis (PCA), eliminate irrelevant features through dimensionality reduction, considering the similarity or correlation between data. On the other hand, semi-supervised techniques, such as Denoising Autoencoder (DAE), utilize labeled and unlabeled data to assess the relevance of features, functioning effectively in scenarios where most of the data is labeled [13,14].

There are several techniques for feature selection. Minimum Redundancy Maximum Relevance (mRMR), developed by [16], chooses a subset of features that contain valuable information for a specific task, minimizing redundancy among them and maximizing their relevance for the task in question. In contrast, Maximum Relevance evaluates the individual importance of each characteristic concerning the target variable. Minimum redundancy considers the correlation between characteristics and each other [8,16].

Neighborhood Component Analysis (NCA) is a dimensionality reduction technique developed by [17], which learns a linear transformation of the original data so that the Euclidean distance in the new representation is meaningful for the classification task, transforming the data into a more discriminative feature space.

Principal Component Analysis is a multivariate analysis technique developed by [18] designed to simplify complex datasets by reducing their dimensionality. The fundamental concept of PCA involves identifying the primary direction of variability within the data and representing it as a single principal component. Subsequently, PCA identifies additional orthogonal directions, each capturing subsequent levels of variability, creating subsequent principal components. This iterative process continues until the principal components encapsulate all significant variations in the data [18–20].

A Denoising Autoencoder (DAE) is an Artificial Neural Network used for unsupervised learning to acquire efficient data representations. Developed by [21], the technique aims to reduce the dimensionality of data by imposing a reconstruction constraint. To achieve this, DAE employs both an encoder and a decoder: the encoder compresses the input data while the decoder reconstructs it. During training, the DAE learns to map noisy

input back to its original form, effectively capturing essential data features while filtering out noise. This process results in a robust network capable of handling noisy data [21].

Table 1 shows some techniques for feature selection, including the purpose and category (filter, wrapper, and embedded) to which each one belongs. In the filter category, there are methods such as Pearson correlation, which measure the linear relationship between variables and help to identify those with the most significant influence on the result. In the wrapper category, there are more sophisticated techniques, such as mRMR, Recursive Feature Elimination (RFE), and Greedy Forward Selection (GFS), which evaluate subsets of variables to find the best combination of features. Finally, Lasso Regression, DAE, and PCA stand out in the embedded category, incorporating feature selection directly into the model training process and adjusting variable coefficients to maximize model performance.

Table 1. Summary of some feature selection techniques found in the literature, their purposes and categories.

Category	Techniques	Purpose
Filter	Pearson Correlation	Quantifies the linear relationship between two variables and can aid in selecting features highly correlated with the target variable [18].
Wrapper	Minimum Redundancy Maximum Relevance	Evaluates each feature's relevance to the target variable while also seeking to minimize redundancy between the selected features [10,16].
	Recursive Feature Elimination	Removes less important features based on the importance assigned by a learning model [14].
	Greedy Forward Selection	Starts with an empty set of features and iteratively adds the best feature that improves model performance [22].
Embedded	Lasso Regression	Incorporates an L1 penalty into the model cost function, leading to automatic feature selection [13].
	Denosing Autoencoder	Reconstructs the input data and converts them into its output. During this process, it can learn compact and informative representations of the data, effectively performing implicit feature selection [21].
	Neighborhood Component Analysis	Creates a projection matrix to maximize the probability of correctly classifying data, which results in the automatic selection of the most discriminative features during model training [11,17].
	Principal Component Analysis	Performs dimensionality reduction by projecting the data into a new feature space composed of the most significant principal components [18–20].

As one can observe, there is a concern regarding feature selection aimed at verifying the identification of the most relevant characteristics, simplifying complexity, and enhancing the computational efficiency of the model. This process accelerates the model training process and improves its generalization ability, reducing the risk of overfitting. Industries can identify which characteristics of the collected data are most pertinent to the problem under study, enabling them to prioritize the deployment of sensors that capture these specific features, thus ensuring that the collected data are as informative as possible for machine learning models.

2.2. Machine Learning Techniques Used to Predict Industrial Machine Failures

The advancement of high-performance technologies has resulted in an exponential increase in the amount of data collected, primarily through the implementation of Internet of Things (IoT) devices, in which sensors play a crucial role in capturing information from the environment. They have been fundamental for collecting data in several areas, such as industrial manufacturing, transportation and mobility, energy, retail, smart cities, health, the supply chain, and agriculture, among others [1,8,23]. Therefore, these devices bring significant opportunities for companies; the acquisition of IoT devices and the choice of appropriate sensors require careful analysis and a strategic approach to ensure success and the expected return on investment [1,8,23].

In recent years, several studies have tackled the challenge of predicting and diagnosing industry failures, underscoring the pivotal role of advanced techniques like machine learning and neural networks. The study of [24] innovates with their approach to diagnosing bearing failures, introducing the novel Logistic-ELM (Extreme Learning Machine). In [9,25], the authors developed models to estimate the state of rotating components. The study of [25] used an SVM-based system, and the authors of [9] used SVM and feature selection applying ReliefF and PCA, highlighting the applicability of machine learning and feature selection methods in this scenario. In [11], the authors proposed a hybrid approach for fault classification in power transmission networks. They employed feature selection using NCA, adding complexity to diagnostic strategies. In [26], the authors created a model to predict failure in beam–column junctions, employing machine learning techniques including K-Nearest Neighbor (KNN), Linear Regression (LR), Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), Extra Tree (ET), AdaBoost (AB), Light Gradient Boosting Machines (GBDTs), and Extreme Gradient Boost (XGBoost). This broad spectrum of techniques showcases the adaptability of machine learning methods. This research has also extended to other areas, such as predicting failures of the high-pressure fuel system [10] and failure prediction in forced blowers [27]. The work [10] applied LR, RF, XGBoost, SVM, and MLP NN with feature selection using mRMR, while in [27], LR, SVM, KNN, XGBoost, and RF techniques were used without feature selection.

A highlight is the work of [28], who applied joint reserve intelligence and feature selection techniques Classifier Attribute Evaluation (CAE), Correlation Attribute Evaluation (COAE), Infogain Subset Evaluation (ISE), and Classifier Subset Evaluation (CSE) to predict machine failures. This indicates a trend towards incorporating innovative approaches to improve predictive effectiveness. In [29], the authors developed an approach based on an ensemble of convolution-based methods for fault detection using vibration signals.

Table 2 summarizes machine learning techniques for predicting machine failure, including the objective of each work, the machine learning techniques used, and the performance metrics used to evaluate the models.

Table 2. Summary of techniques, objectives, and metrics employed by some works addressing machine failure prediction.

Work	Main Objective	Techniques Employed	Performance Evaluation Metrics
[6]	Failure prediction in tube filling machine	RF, LR	Accuracy, MSE
[9]	Fault diagnosis in rotating bearings	PCA, SVM	Accuracy
[10]	Predict high-pressure fuel system failures based on machine learning techniques and a multilayer neural network	mRMR, LR, RF, XGBoost, SVM, MLP	Precision, Recall, F-Measure
[11]	Classification of failures in transmission networks	NCA, SVM, KNN, MLP	Accuracy
[24]	Diagnosis of bearing failures using Logistic-ELM	LELM	Accuracy
[25]	Estimate the state of rotating components	SVM-based system	Accuracy
[26]	Prediction of failure modes in beam–column joints	KNN, LR, SVM, ANN, DT, RF, ET, AB, GBDT, XGBoost	Accuracy, Precision, Recall, and AUC
[27]	Prediction of failures in forced blowers	LR, SV, KNN, XGBoost, RF	Matthew’s correlation coefficient (MCC) and AUC
[28]	Predicting machine failures with joint reserve intelligence	DT, BN, LR, JRI	Accuracy, Precision, Recall, F-Measure, MAE
[29]	Multivariate time series of vibration signals of a machine	Ensemble, LSTM, Mini Rocket, Resnet	Accuracy
[30]	Resource selection and SVM for maintenance failure prediction	RF, SVM, SMOTE	Accuracy

The literature on machine failure prediction covers a wide variety of approaches and techniques. Many studies focus on developing machine learning algorithms [9–11,24–30]. Others explore methods of processing signals acquired from machines through sensors [25,31]. There is also research investigating the use of image processing techniques to identify visual anomalies in mechanical components. These studies aim to improve the efficiency, reliability, and safety of industrial operations, contributing to predictive maintenance and reducing unplanned downtime.

3. Methodology

This section describes the methodology employed to achieve the objectives outlined in this study, covering data collection and normalization, a comparison of four feature selection techniques, the application of three machine learning techniques to develop prediction models, model evaluation and comparison, and the classification of machine failures by using the best model. Detailed descriptions of these steps, illustrated in Figure 1, are presented below.

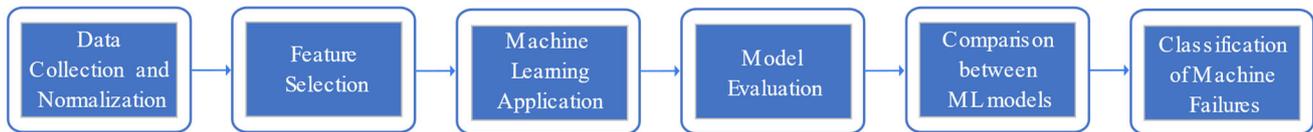


Figure 1. Steps carried out in this work.

3.1. Data Collection and Normalization (Step 1)

This step involves collecting and pre-processing data. First, a dataset related to predictive maintenance, proposed by [32] and adopted in the study of [28], was collected from the UC Irvine Machine Learning Repository. This dataset has 10,000 records, each composed of the following attributes: UID, a unique identifier ranging from 1 to 10,000; ProductID, which represents the quality of the product, where L represents 50% of all products, M represents 30% of all products, and H represents 20% of all products; air temperature (K), generated using a random walk process subsequently normalized to a standard deviation of 2 K around 300 K; process temperature (K), generated using a random walk process normalized to a standard deviation of 1 K, added to the air temperature plus 10 K; rotation speed (rpm), calculated from the power of 2860 W, overlaid with customarily distributed noise; torque (Nm) values are typically distributed around 40 Nm with $\sigma = 10$ Nm and no negative values; tool wear (min), H/M/L quality variants add 5/3/2 min of tool wear to the tool used in the process; target label indicates whether the machine failed or not at this specific data point; and failure type represents a specific type of failure, including no failure, heat dissipation failure, power failure, random failure, overstrain failure, and tool wear failure.

The data are imbalanced, with a low incidence of some types of failures. There are 9652 no failure samples, 112 heat dissipation failure, 95 power failure samples, 78 overstrain failure samples, 45 tool wear failure samples, and 18 random failure samples.

In the pre-processing procedure, the data were normalized to the interval [0, 1] using the Min-Max approach with the resources of the MinMaxScaler library from Sklearn in Python to avoid differences in the scales of the variables.

Since real predictive maintenance data are often difficult to obtain, this dataset represents an important source of data, because according to [28,32], it reflects the real predictive maintenance found in the industry.

3.2. Feature Selection and Comparison of Techniques (Step 2)

This step involves applying and comparing the techniques Principal Component Analysis (PCA), Minimum Redundancy Maximum Relevance (mRMR), Neighborhood Component Analysis (NCA), and Denoising Autoencoder (DAE) to select the best features related to machine failures.

In this work, PCA was applied to the data as indicated by [19,20] to verify each variable's percentage of total variance using the main components. To do this, we employed Sklearn's PCA Decomposition library in Python, which has a function called `explained_variance_ratio`. This function calculates the cumulative proportion of the explained variance and returns the total number of principal components with the percentage contribution of each variable to each component. The mRMR was provided by `pymrmr` library in Python. It receives the input and output variables and returns the relevance level of each input variable about the output. The `sklearn.neighbors` library was used to provide

the correlation levels between the input and output variables using NCA. The procedure involves showcasing input and output variables while partitioning 70% of the data for training and reserving 30% for testing. Finally, the DAE technique was applied using Keras and Tensorflow version 2.15.0. As mentioned before, it is a useful tool to deal with dimensionality reduction and feature selection.

In this study, we employ the technique using TensorFlow's input, model, and dense libraries. Keras.layers and models. We employ a dense with nine inputs and relu activation function for the encoding layer, while we use the linear activation function for decoding. Furthermore, we applied the Adam optimizer with 100 epochs, a batch size 64, and the mean squared error metric as a loss function. To evaluate the contributions of each variable, we extract the weights from the autoencoder-encoded layer and apply them to the corresponding variables.

Each of these techniques was employed to extract important features from the data and reduce the dimensionality of the feature set. We also sought to observe the level of importance attributed to each resource through these techniques. Subsequently, we compared the best features highlighted by each technique.

It is important to highlight that the RF technique, being based on Decision Trees (DTs), can be used for feature selection by analyzing the contribution of each attribute (variable) to the model's prediction capability. During the training of the RF algorithm, attributes are evaluated in each individual DT to measure how much each one contributes to reducing impurity in the tree nodes. Then, the importance of each attribute is calculated by the average or sum of impurity reductions across all DTs in the forest. Thus, attributes with greater impurity reductions are considered more important for the model, while those with lesser impact are deemed less relevant and may even be discarded during the DT construction process. Therefore, we also included in the results an analysis of the importance of attributes based on RF.

3.3. Failure Classification Using Machine Learning Techniques (Step 3)

Given its effectiveness, as pointed out by researchers in [33], we developed a Random Forest (RF) model to classify the failure type. RF is a machine learning technique that creates an ensemble of decision trees during training and then makes predictions based on decisions from those trees. For classification, each tree "votes" a class for a specific input [6,10,27,33].

Figure 2 illustrates the process of RF training, which considers the most important input variables, defined by the feature selection technique, to produce the output indicating the type of failure after each tree votes for one of the six possible failure classes. The model training was conducted with 50, 100, and 200 estimators (trees), separating 70% of the data for training and 30% for testing. We used accuracy, precision, and recall metrics to evaluate the model, as indicated in [11,26–28].

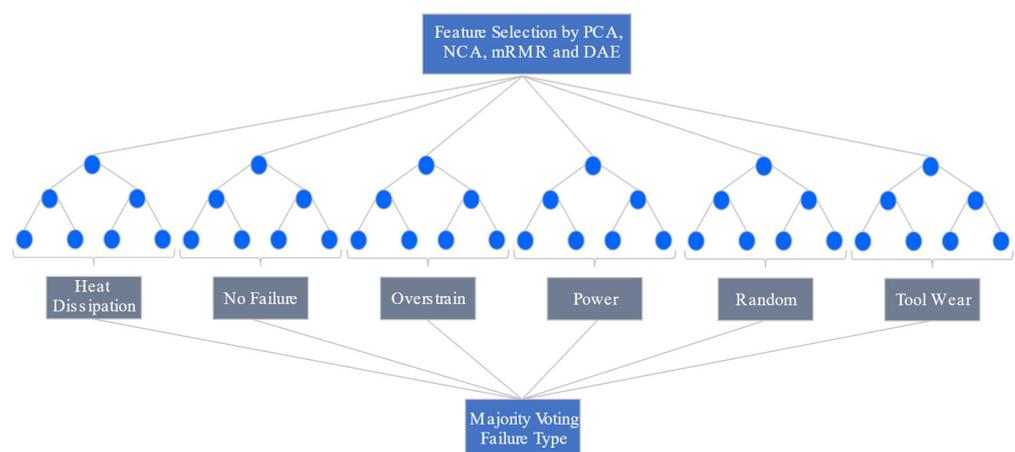


Figure 2. Failure classification using Random Forest.

Afterward, the Support Vector Machine (SVM) technique was employed. Its objective is to maximize the margin between different planes. SVM seeks to find a hyperplane that separates instances of different classes in a feature space. A hyperplane is the decision surface that maximizes the margin (support vectors) between opposing class examples [3,10,20].

Figure 3 illustrates the architecture of the SVM model for failure classification, which comprises input, hidden, and output layers. In the hidden layer, the linear kernel functions, regularization equal to 1, and gamma equal to scale were used.

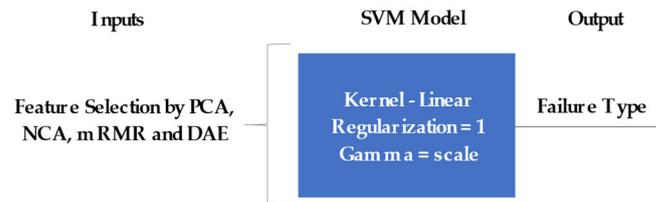


Figure 3. Failure classification using SVM.

Finally, the Multilayer Perceptron Neural Network (MLP) technique was used to classify machine failures. This sophisticated architecture includes an input layer for data reception, hidden layers for applying non-linear transformations via activation functions, and an output layer for generating predictions. Each neuron connection has a weight, and the output layer’s function depends on the task, which can be sigmoid for binary classification or SoftMax for multiclass classification. The training process involves forward propagation for network output calculation, a loss function for output comparison, and backpropagation for weight and bias adjustment. Optimization algorithms iterate this process to minimize loss [7,26,34].

Figure 4 illustrates the MLP neural network’s architecture used in this research. It comprises three layers: input, hidden, and output. The number of neurons in the hidden layer (n_h) was determined using the method developed by [34], described by Equation (1), in which n_i and n_o are, respectively, the number of input and output variables.

$$n_h = n_o \left(\sqrt[3]{\frac{n_i}{n_o}} \right)^2 \tag{1}$$

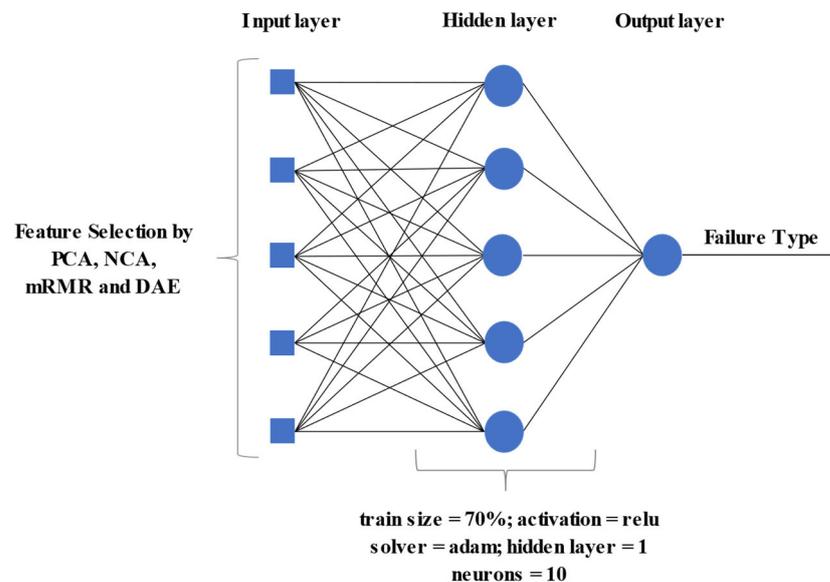


Figure 4. Failure classification using MLP.

Additionally, the parameters adopted include the SoftMax activation function and the Adam solver, which are most used in the literature, as indicated by [34]. The training was conducted with 100 epochs using 70% of the data. The remaining 30% was used for testing.

3.4. Model Evaluation (Step 4)

Accuracy, precision, and recall metrics were utilized to evaluate the performance of RF, SVM, and MLP. The accuracy calculates the proportion of correctly classified instances among the total number of instances in the dataset. Precision is a metric that measures the proportion of correctly predicted positive cases (true positives) among all instances predicted as positive, regardless of whether they were actually positive or negative. The recall metric focuses on the proportion of actual positive cases that are correctly identified by the model and is calculated as the ratio of true positives to the sum of true positives and false negatives [10,26].

3.5. Model Comparison (Step 5)

In this step, the three classification models, RF, SVM, and MLP, are compared to determine which machine learning technique offers the best performance in machine failure classification. Performance metrics such as accuracy, precision, and recall are used to analyze each model's capabilities comprehensively.

Finally, step 6 selects the model that performs best according to these metrics to classify machine failures.

4. Results and Discussions

First, the results are presented by applying feature selection techniques (PCA, mRMR, NCA, and DAE). Then, considering the most relevant features, we evaluate the performance of RF, SVM, and MLP in predicting machine failures in terms of accuracy, precision, and recall. In both cases, feature selection and classification, the results obtained by the techniques are compared.

Table 3 shows how each feature (variable) contributes to every principal component. The table allows for observing the nine main components, PC1 to PC9, and the variables contributing to each principal component. Product ID contributes 100% to PC1, while air temperature and process temperature contribute 73.4% to PC2, followed by rotational speed at 100% and torque at 98.7%. It means that with just two main components, it is possible to explain the proportion of the total variance of the data, showing that the variables product ID, air temperature, process temperature, rotation speed, and torque are most important to predict machine failure. Although the first two components explain most of the variance (about 98%), the remaining components can provide important information about less prominent patterns in the data. Therefore, presenting all components allows for a more complete and detailed view of the underlying structure of the data.

Table 3. Contribution of each variable to each principal component.

Variables	Contribution (in %)								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Product ID	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Air Temperature	0.0	73.4	100.0	2.0	3.1	100.0	11.1	42.9	7.1
Process Temperature	0.0	73.4	98.7	2.0	3.1	95.7	11.1	47.6	7.1
Rotation speed	0.0	100.0	74.0	0.0	53.1	13.0	100.0	0.0	7.1
Torque	0.0	98.7	74.0	0.0	53.1	13.0	100.0	0.0	7.1
Tool Wear	0.0	2.5	3.9	100.0	6.3	0.0	5.6	0.0	0.0

Table 4 presents the feature selection results for PCA, mRMR, NCA, and DEA techniques. The mRMR presents product ID and tool wear as the most relevant variables, while the NCA presents practically the same contribution level to all input variables. On the other hand, PCA presents the five most important variables: product ID, rotational speed,

torque, air temperature, and process temperature. When comparing the results of the three techniques, only the product ID variable has a similar contribution level using both mRMR (26.56%) and PCA (22.07%), showing that this variable can be used in the prediction model. However, the mRMR also presents a high level of correlation for the variable tool wear (10.95%). The DAE revealed that air temperature, process temperature, product ID, and tool wear are the four most relevant variables for the failure prediction process.

Table 4. Importance of features according to each feature selection technique.

Features	PCA	mRMR	NCA	DAE
Product ID	22.07	26.56	12.50	11.33
Air temperature	16.20	1.27	12.40	12.41
Process temperature	16.20	0.78	12.40	12.33
Rotational speed	22.07	6.68	12.30	10.78
Torque	21.78	7.22	12.70	10.03
Tool wear	0.55	10.95	12.90	11.31

As pointed out in [10], feature selection provides better performance for models to predict machine failures, achieving an accuracy of 82.4% using mRMR. Regarding the use of NCA for feature selection, although [11] presents a gain of 5 to 10% in the accuracy of fault classification by applying this technique, in our study, it was not possible to achieve this gain, as this technique presented practically the same level of importance for all features, as can be seen in Table 4.

It is worth noting that the RF defined the following order of importance of attributes: torque, rotational speed, tool wear, product ID, air temperature, and process temperature. This order is different from those made by the three attribute selection techniques, which prompts further investigation into the effectiveness of the impurity metrics employed by RF to define the importance of attributes to failure classification.

Table 5 shows the results obtained by RF, SVM, and MLP models on training and test data, accompanied by the training time for each model. In the case of RF considering feature selection, we observed good performance in the training and test data, maintaining high accuracy, precision, and recall rates, which were 1.0 for the three metrics in the training and 0.98, 0.97, and 0.98 for the test data. The SVM performed well, but it required extensive training time compared to RF and MLP, and despite the MLP demonstrating robustness in training and testing data, RF proved to be more efficient.

Table 5. Models' performance for training and testing data.

Model	Train Data			Test Data			Training Time (s)	Feature Selection
	Accuracy	Precision	Recall	Accuracy	Precision	Recall		
RF	1.0	1.0	1.0	0.98	0.97	0.98	25	Yes
SVM	0.97	0.96	0.97	0.98	0.96	0.98	331	
MLP	0.98	0.96	0.98	0.98	0.97	0.98	2	
RF	0.98	0.98	0.98	0.95	0.94	0.95	3	No
SVM	0.96	0.94	0.96	0.95	0.94	0.96	593	
MLP	0.96	0.94	0.96	0.96	0.94	0.96	6	

Without the application of feature selection, RF was more efficient than the other two classifiers, but with lower rates for the three evaluation metrics in the tests. The SVM showed promising results but with a significantly longer training time, and the MLP preserved solid performance.

Overall, the models obtained promising training and testing data results without using strategies to mitigate overfitting, such as cross-validation and regularization.

In Table 6, precision and recall metrics are provided for different failures predicted by the three models analyzed. It is noteworthy that the RF model demonstrates solid performance in failure prediction, especially for heat dissipation failure (85%), no failure (98%), power failure (86%), and overstrain failure (100%), indicating its high performance in identifying these types of failure. In contrast, the SVM model exhibits varied performance for different failure categories, presenting good accuracy for no failure (100%), power failure (81%), and heat dissipation failure (81%), but facing difficulties in detecting failures overstrain, random, and toolwear, for which a rate of 0% was obtained for precision and recall.

Table 6. Models' performance for each failure type.

Failure Type	RF		SVM		MLP	
	Precision	Recall	Precision	Recall	Precision	Recall
Heat Dissipation Failure	0.85	0.87	0.73	0.35	1.00	0.57
No Failure	0.98	1.00	0.98	1.00	0.98	1.00
Overstrain Failure	1.00	0.05	0.00	0.00	0.00	0.00
Power Failure	0.86	0.81	0.96	0.84	0.96	0.77
Random Failure	0.00	0.00	0.00	0.00	0.00	0.00
Toolwear	0.00	0.00	0.00	0.00	0.00	0.00

The MLP model stands out for its solid performance. It achieved a precision of 98% for no failure, 96% for power failure, and 100% for heat dissipation failure. However, this model also faces challenges in classifying overstrain, random, and toolwear failures. Thus, the RF model shows superior overall performance. It obtains high accuracy, precision, and recall for four failure types. However, it failed to classify the random and toolwear failures.

Given these results, the importance of adopting a comprehensive approach when choosing feature selection techniques to analyze datasets is evident, as exemplified by the study of [28] and the results presented in Tables 4 and 5. The careful choice of feature selection technique directly impacts the selection of the most relevant variables, reflecting the model's accuracy as in [10,11] to better support decision-making aimed at predictive maintenance.

Regarding the overall performance of the techniques shown in Tables 4 and 5, RF presented more consistent results in precision and recall than the other techniques in several failure categories, especially in heat dissipation and power failure, in which RF achieved a precision of 0.85 and 0.86, respectively, and a recall of 0.87 and 0.81. RF has demonstrated more robustness than SVM and MLP, as evidenced by its ability to handle cases like no failure, which maintains a high precision of 0.98 and recall of 1.00. However, it is important to emphasize that the fact that the data is imbalanced indeed constitutes one of the reasons for the models to fail in classifying random and tool wear failures.

With respect to attribute selection, although the works of Giordano et al. [10], Chang et al. [11], and Bezerra et al. [19] have addressed industrial applications, none of them address the application of attribute selection techniques to prioritize the selection of the most relevant sensors for data collection.

Research like this can assist in selecting industrial sensors, as some may be acquired and used without significantly contributing to detecting or preventing machine failures. This is because not all sensors are equally relevant for monitoring or predicting machine failures. In the example addressed in this research, a company could only use sensors that measure air and process temperature, rotation speed, and torque while optionally monitoring tool wear due to this variable not significantly contributing to failure prediction. Thus, the framework presented here could result in a cost reduction associated with sensor acquisition.

5. Conclusions

This study presents a framework based on machine learning techniques for failure machine prediction, with a particular emphasis on feature selection methodologies. By addressing a tangible industrial challenge that impacts production planning, this research offers a practical solution and delves into the importance of selecting relevant features for accurate predictions.

Feature selection techniques, including mRMR and PCA, highlight crucial variables, shedding light on the most influential factors in predicting machine failures. Notably, the PCA method, emphasizing variables like product ID, air and process temperature, rotation speed, and torque, was instrumental in enhancing the performance of the RF prediction model, which demonstrated itself to be superior to SVM and MLP.

The examples addressed in this work suggest that decision-makers in industries are highly recommended to invest in thermocouple sensors and thermistors for measuring air and process temperature, tachometers and magnetic encoders for measuring rotational speed, and torque sensors with magnetic and piezoelectric effects. These variables have been proven to be more significant in collecting information about torque.

The example covered in this work suggests that decision-makers in industries invest in thermocouple sensors and thermistors for measuring air and process temperatures, tachometers and magnetic encoders for measuring rotational speed, and torque sensors with magnetic and piezoelectric effects. This is because these variables proved to be more significant in collecting information about the machines.

In modern industrial environments, where efficiency and reliability are fundamental, investing in predictive machine failure models is essential to sustaining competitiveness and operational longevity. Furthermore, the insights gleaned from this research raise discussion regarding the significance of sensor selection, guiding industries to prioritize those that capture data that leads to machine failure predictions more effectively. This holistic approach strengthens predictive maintenance strategies and sustains operational excellence in dynamic industrial scenarios.

In future work, we intend to address the problem of data imbalance by applying resampling techniques, investigate the use of metaheuristic approaches to optimize data acquisition, explore in more depth the selection of attributes by the RF algorithm itself using impurity metrics, such as the Gini index and entropy, and apply model interpretability techniques to better understand how machine learning techniques predict machine failures.

Author Contributions: Conceptualization, F.E.B., G.C.d.O.N., G.M.C., R.F.M., A.M.d.F. and M.V.; Methodology, F.E.B., G.C.d.O.N., G.M.C., R.F.M., A.M.d.F. and M.V.; software, F.E.B.; Validation, F.E.B., G.C.d.O.N., G.M.C., R.F.M., A.M.d.F. and M.V.; Formal analysis, F.E.B., G.C.d.O.N., S.A.d.A. and G.A.L.; Investigation, F.E.B. and G.C.d.O.N.; Resources, F.E.B. and G.C.d.O.N.; Data curation, F.E.B.; Writing—original draft preparation, F.E.B., G.C.d.O.N. and S.A.d.A.; Writing—review and editing, F.E.B., G.C.d.O.N. and S.A.d.A. and M.A.; Visualization, F.E.B., G.C.d.O.N., S.A.d.A., M.A. and M.S.; Supervision, G.C.d.O.N. and M.A.; Project administration, G.C.d.O.N. and M.A.; Funding acquisition, M.S. and M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the research unit on Governance, Competitiveness and Public Policy (UIDB/04058/2020) and (UIDP/04058/2020), funded by national funds through FCT—Fundação para a Ciência e a Tecnologia. The authors are grateful to CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico—Research funding in Productivity (PQ-2 09/2023-Process: 305950/2023-1).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://archive.ics.uci.edu/dataset/601/ai4i+2020+predictive+maintenance+dataset> (accessed on 12 January 2024).

Acknowledgments: The authors would like to thank University of São Paulo (USP), Federal University of ABC, FEI University, and University of Aveiro for all the support provided throughout the research.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

AB	AdaBoost
CAE	Classifier Attribute Assessment
COAE	Correlation Attribute Assessment
CSE	Classifier Subset Evaluation
DAE	Denoising Auto Encoder
DL	Deep Learning
DT	Decision Tree
ELM	Extreme Learning Machine
ET	Extra Tree
GB	Gradient Boosting
GBT	Gradient Boosted Tree
GFS	Greedy Forward Selection
ISE	Infogain Subset Valuation
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
LR	Linear Regression
MAE	Mean Absolute Error
MCC	Matthews Correlation Coefficient
MLP	Multilayer Perceptron
mRMR	Minimum Redundancy Maximum Relevance
NCA	Neighborhood Component Analysis
PCA	Principal Component Analysis
R2	Coefficient of Determination
RF	Random Forest
RFE	Recursive Feature Elimination
RMS	Root Mean Square
RMSE	Root Mean Square Error
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting

References

1. Javaid, M.; Haleem, A.; Singh, R.P.; Rab, S.; Suman, R. Significance of sensors for industry 4.0: Roles, capabilities, and applications. *Sens. Int.* **2021**, *2*, 100110. [[CrossRef](#)]
2. Kwon, O.; Sim, J.M. Effects of data set features on the performances of classification algorithms. *Expert Syst. Appl.* **2013**, *40*, 1847–1857. [[CrossRef](#)]
3. Yafooz, W.; Bakar, Z.; Fahad, S.; Mithun, A. Business Intelligence Through Big Data Analytics, Data Mining and Machine Learning. *Adv. Intell. Syst. Comput.* **2020**, *1016*, 17–33. [[CrossRef](#)]
4. Zonta, T.; Da Costa, C.A.; da Rosa Righi, R.; de Lima, M.J.; da Trindade, E.S.; Li, G.P. Predictive maintenance in the Industry 4.0: A systematic literature review. *Comput. Ind. Eng.* **2020**, *150*, 106889. [[CrossRef](#)]
5. Pech, M.; Vrchota, J.; Bednář, J. Predictive Maintenance and Intelligent Sensors in Smart Factory: Review. *Sensors* **2021**, *21*, 1470. [[CrossRef](#)] [[PubMed](#)]
6. Natanael, D.; Sutanto, H. Machine Learning Application Using Cost-Effective Components for Predictive Maintenance in Industry: A Tube Filling Machine Case Study. *J. Manuf. Mater. Process.* **2022**, *6*, 108. [[CrossRef](#)]
7. Mateo Casali, M.A.; Fraile Gil, F.; Boza, A.; Nazarenko, A. An industry maturity model for implementing Machine Learning operations in manufacturing. *Int. J. Prod. Manag. Eng.* **2023**, *11*, 179–186. [[CrossRef](#)]
8. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. In *Data Classification: Algorithms and Applications*; CRC Press: Boca Raton, FL, USA, 2014; pp. 37–64. [[CrossRef](#)]
9. Rajeswari, C.; Sathiyabhama, B.; Devendiran, S.; Manivannan, K. Bearing fault diagnosis using multiclass support vector machine with efficient feature selection methods. *Int. J. Mech. Mechatron. Eng.* **2015**, *15*, 1–12.
10. Giordano, D.; Pastor, E.; Giobergia, F.; Cerquitelli, T.; Baralis, E.; Mellia, M.; Tricarico, D. Dissecting a data-driven prognostic pipeline: A powertrain use case. *Expert Syst. Appl.* **2021**, *180*, 115109. [[CrossRef](#)]

11. Chang, G.W.; Hong, Y.H.; Li, G.Y. A hybrid intelligent approach for classification of incipient faults in transmission network. *IEEE Trans. Power Deliv.* **2019**, *34*, 1785–1794. [[CrossRef](#)]
12. Jemai, J.; Zarrad, A. Feature Selection Engineering for Credit Risk Assessment in Retail Banking. *Information* **2023**, *14*, 200. [[CrossRef](#)]
13. Okech, E.O.; Okeyo, G.O.; Kimwele, M.W. Feature Selection for Classification using Principal Component Analysis and Information Gain. *Expert Syst. Appl.* **2021**, *174*, 114765. [[CrossRef](#)]
14. Saeys, Y.; Inza, I.; Larranaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [[CrossRef](#)] [[PubMed](#)]
15. Khalid, S.; Khalil, T.; Nasreen, S. A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning. In Proceedings of the 2014 Science and Information Conference, London, UK, 27–29 August 2014; pp. 372–378. [[CrossRef](#)]
16. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)] [[PubMed](#)]
17. Kariuki, H.; Mwalili, S.; Waititu, A. Dimensionality Reduction of Data with Neighbourhood Components Analysis. *Int. J. Data Sci. Anal.* **2022**, *8*, 72–81. [[CrossRef](#)]
18. Pearson, K. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
19. Bezerra, F.E.; Grassi, F.; Dias, C.G.; Pereira, F.H. A PCA-based variable ranking and selection approach for electric energy load forecasting. *Int. J. Energy Sect. Manag.* **2022**, *16*, 1172–1191. [[CrossRef](#)]
20. Schimit, P.H.; Pereira, F.H. Disease spreading in complex networks: A numerical study with principal component analysis. *Expert Syst. Appl.* **2018**, *97*, 41–50. [[CrossRef](#)]
21. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.-A. Extracting and Composing Robust Features with Denoising Autoencoders. In Proceedings of the 25th International Conference on Machine Learning (ICML'08), Helsinki, Finland, 5–9 July 2008; Association for Computing Machinery: New York, NY, USA, 2008; pp. 1096–1103. [[CrossRef](#)]
22. Lei, Y.; Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003.
23. Mouha, R. Internet of Things (IoT). *J. Data Anal. Inf. Process.* **2021**, *9*, 77–101. [[CrossRef](#)]
24. Tan, Z.; Ning, J.; Peng, K.; Xia, Z.; Wu, D. Logistic-ELM: A novel fault diagnosis method for rolling bearings. *J. Braz. Soc. Mech. Sci. Eng.* **2022**, *44*, 553. [[CrossRef](#)]
25. Ruiz-Gonzalez, R.; Gomez-Gil, J.; Gomez-Gil, F.J.; Martínez-Martínez, V. An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis. *Sensors* **2014**, *14*, 20713–20735. [[CrossRef](#)]
26. Gao, X.; Lin, C. Prediction model of the failure mode of beam-column joints using machine learning methods. *Eng. Fail. Anal.* **2020**, *120*, 105072. [[CrossRef](#)]
27. Salem, K.; AbdelGwad, E.; Kouta, H. Predicting Forced Blower Failures Using Machine Learning Algorithms and Vibration Data for Effective Maintenance Strategies. *J. Fail. Anal. Preven.* **2023**, *23*, 2191–2203. [[CrossRef](#)]
28. Shaheen, A.; Hammad, M.; Elmedany, W.; Ksantini, R.; Sharif, S. Machine failure prediction using joint reserve intelligence with feature selection technique. *Int. J. Comput. Appl.* **2023**, *45*, 638–646. [[CrossRef](#)]
29. Lee, X.Y.; Kumar, A.; Vidyaratne, L.; Rao, A.R.; Farahat, A.; Gupta, C. An ensemble of convolution-based methods for fault detection using vibration signals. In Proceedings of the 2023 IEEE International Conference on Prognostics and Health Management (ICPHM), Montreal, QC, Canada, 5–7 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 172–179. [[CrossRef](#)]
30. Tarik, M.; Mniai, A.; Jebari, K. Hybrid feature selection and support vector machine framework for predicting maintenance failures. *Appl. Comput. Sci.* **2023**, *19*, 112–124. [[CrossRef](#)]
31. Ogaili, A.A.F.; Jaber, A.A.; Hamzah, M.N. A methodological approach for detecting multiple faults in wind turbine blades based on vibration signals and machine learning. *Curved Layer. Struct.* **2023**, *10*, 20220214. [[CrossRef](#)]
32. Matzka, S. Explainable artificial intelligence for predictive maintenance applications. In Proceedings of the 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), Irvine, CA, USA, 21–23 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 69–74. [[CrossRef](#)]
33. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
34. Rachmatullah, M.I.C.; Santoso, J.; Surendro, K.A. Novel Approach in Determining Neural Networks Architecture to Classify Data with Large Number of Attributes. *IEEE Access* **2020**, *8*, 204728–204743. [[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.