

## Article

# Prediction of Maintenance Activities Using Generalized Sequential Pattern and Association Rules in Data Mining

Abbas Al-Refaie <sup>1</sup>, Banan Abu Hamdiah <sup>1</sup> and Natalija Lepkova <sup>2,\*</sup><sup>1</sup> Department of Industrial Engineering, The University of Jordan, Amman 11942, Jordan<sup>2</sup> Department of Construction Management and Real Estate, Faculty of Civil Engineering, Vilnius Gediminas Technical University, Sauletekio Av. 11, 10223 Vilnius, Lithuania

\* Correspondence: natalija.lepkova@vilniustech.lt

**Abstract:** This study proposed a data mining framework for predicting sequential patterns of maintenance activities. The framework consisted of data collection, prediction of maintenance activities with and without attributes, and then the comparison between prediction results. In data collection, historical data were collected regarding maintenance activities and product attributes. The generalized sequential pattern (GSP) and association rules were then applied to predict maintenance activities with and without attributes to determine the frequent sequential patterns and significant rules of maintenance activities. Finally, a comparison was performed between the sequences of maintenance activities with and without attributes. A real case study of washing machine products was presented to illustrate the developed framework. The results showed that the proposed framework effectively predicted the next maintenance activities and planning preventive maintenance based on product attributes. In conclusion, the data mining approach is found effective in determining the maintenance sequence that reduces downtime and thereby enhancing productivity and availability.

**Keywords:** prediction of maintenance; data mining; generalized sequential pattern; association rule mining; maintenance planning



**Citation:** Al-Refaie, A.; Abu Hamdiah, B.; Lepkova, N. Prediction of Maintenance Activities Using Generalized Sequential Pattern and Association Rules in Data Mining. *Buildings* **2023**, *13*, 946. <https://doi.org/10.3390/buildings13040946>

Academic Editor: Chyi Lin Lee

Received: 27 February 2023

Revised: 25 March 2023

Accepted: 29 March 2023

Published: 3 April 2023



**Copyright:** © 2023 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

Maintenance is treated as one of the critical functions in any organization [1–3]. Facility maintenance and operations contain two main task areas including (1) the care and maintenance of the building and technical systems (including the repair and replacement of technical systems, and equipment) and (2) the care and maintenance of outdoor areas [4]. In practice, an unexpected failure reduces production, decreases efficiency, increases cost, and affects personal safety [5,6]. A random failure of an equipment part can lead to another failure in other parts of the system [7,8]. Hence, maintenance prediction (MP) has become a crucial maintenance strategy that aims to avoid the occurrence of failure and reduce the downtime of machines and maintenance costs.

MP aims at forecasting whether the failure of a component in a system has occurred at a specific point of time in the future. Predictive Maintenance (PdM) is usually performed to minimize preventive and corrective maintenance [9]. A PdM system is described as a set of components that measure one or more input variables and processes to estimate the failure probability of equipment's future state in the system [10,11]. It usually analyses the past and current data that illustrated the system status, events, and operations [12]. Extracting the relationships among transactions is very useful to make a knowledge base for predicting failure and maintenance requirements, as well as analysing possible causes for the deviations [13]. Therefore, this research proposes a framework for maintenance prediction of the future sequence of maintenance activities (repair or replacement) of faulty product parts.

The data mining (DM) approach is reported to be effective in extracting the relationship between failures and predicting the failure using events sequences of historical data [14,15].

DM is defined as a field that combined computer science and statistics used to extract hidden information and useful knowledge and find the correlation between events from large data [16]. It can be classified into two categories: descriptive and predictive. Descriptive is a technique that describes the input data. Predictive is a technique that is implemented on the input to predict the required output [17]. DM techniques can be performed to reduce downtime and related costs by determining the occurrence of the next activity and identifying the dependencies between maintenance activities, and then predicting the occurrence of each maintenance sequence based on historical data [18]. DM techniques can be used for association rule mining, classification, prediction, and clustering [19].

Furthermore, the DM process is classified based on the data type into various categories: sequential pattern mining, association rule mining, classification, and prediction. Sequential pattern mining has many applications in customer behaviour and web log mining. It identifies the pattern of activities among all objects during a specific time interval, while frequent sequential patterns can be determined using the support or threshold value. It can be performed using the Apriori algorithm. Some attributes, such as object number, maintenance type, spare parts used, and maintenance date are included in the transaction to determine a sequential pattern. Association rule mining uses the frequent pattern to identify the relationship between items that occurred in the first period and items that may occur in the next period using the IF-THEN rules. The strength of the association rule is then measured using the support and confidence values. Finally, classification and prediction propose a model to predict specific attributes considering the variability of the data [20].

Data mining techniques have been widely employed in prediction. For example, Jeong et al. [21] developed a decision support model for determining the target multi-family housing complex for green remodeling using data mining techniques. Jeong et al. [22] proposed a data-driven approach for establishing a CO<sub>2</sub> emission benchmark for a multi-family housing complex using data mining techniques. Lv et al. [23] proposed a complex data fusion and efficient learning algorithm (multi-graphics processing unit (GPU)) to process the multi-dimensional and complex big data based on the compositive rough set model. Zhang et al. [24] proposed a novel category-induced coarse-to-fine domain adaptation approach (C2FDA) for cross-domain object detection. Zhang et al. [25] analyzed the problems of existing container positioning methods and proposed a vision-based container position measuring system to provide precise parameters for container lifting operations. Mitici et al. [26] employed dynamic predictive maintenance for multiple components using data-driven probabilistic remaining useful life prognostics illustrated by the case of turbofan engines. Zhou et al. [27] explored the applicability of data mining and knowledge discovery in combination with geographic information system technology to allow management to better decide maintenance strategies, set rehabilitation priorities, and make investment decisions. Mining algorithms including decision trees and association rules were used in the analysis. The selected rules were employed to predict the maintenance and rehabilitation strategy of road segments. A pavement database covering four counties within the state of North Carolina, which was provided by North Carolina DOT, was used to test this method. Dindarloo and Siامي-Irdemoosa [28] investigated the application of classification and clustering approaches for pattern recognition and failure forecasting on mining shovels. The failure behaviour of a fleet of ten mining shovels during one year of operation was examined. The shovels were classified into four clusters using k-means clustering algorithms. Future failures were predicted using the support vector machine classification technique. Historical data for failure and time to repair were used to predict the next failure type for all shovels. Moharana et al. [29] suggested a framework for extracting the sequential patterns of maintenance activities and related spare parts information from historical records of maintenance data with pre-defined support or threshold values. Gharoun et al. [30] proposed an algorithm for fault detection in terms of condition-based maintenance with data mining techniques for an aircraft turbofan engine using flight data. The data-driven models were used to model the relationship between engine exhaust gas

temperature (EGT) and other operational and environmental parameters of the engine. The faults occurring in each flight were detected based on the identification of abnormal events by a one-class support vector machine trained by the health condition EGT residual data set. Gholami and Hafezalkotob [31] combined data mining techniques and time series models to schedule maintenance activities. The clustering algorithm was adopted to categorize failures based on the similarity in types of maintenance activities. Then, rules were extracted for characterizing the clusters and presenting a range for each factor by applying a proper association rule algorithm. Subsequently, time series models were employed to predict the period that a factor may meet its rule's range. Kalathas and Papoutsidakis [32] adopted stored-inactive data from a Greek railway company, used the method of data mining, and applied machine learning techniques to create strategic decision support and develop a risk and control plan for trains. Carrasco et al. [33] generalised the concept of positive and negative instances into intervals to evaluate unsupervised anomaly detection algorithms. They proposed the Preceding Window ROC, a generalisation for the calculation of ROC curves for time series scenarios, and adapted the mechanism from an established time series anomaly detection benchmark to the proposed generalisations to reward early detection. The proposed evaluation method was evaluated by a case study of big data algorithms with a real-world time series problem provided by the ArcelorMittal company.

However, this research utilizes supervised data mining tools, the generation of sequential patterns and association rule mining, for predicting the next maintenance activities for faulty product parts with and without product attributes. The results of this research are valuable to maintenance engineers in the effective planning and management of maintenance activities. A real case study of a washing machine will be provided to illustrate the developed data mining framework.

The remainder of this research including the introduction is structured as follows: Section 2 develops the data mining framework for maintenance prediction. Section 3 presents an application of the framework. Section 4 discusses the research results. Section 5 summarizes research conclusions and future research.

## 2. Development of Data Mining Framework

The developed framework is considered as shown in Figure 1.

Stage 1. Maintenance data regarding maintenance activities and attributes are collected. The types of data may be temporal, such as production date, repair date, and sales date. Categorical data include model and engine type [34]. Numerical data include mileage at repair, maintenance, and spare parts cost. Finally, textual data involve failure description and taken corrective action. For example, assume there are four objects as shown in Table 1, where different maintenance activities (M) are considered during three months. Three maintenance activities are  $M_1$ ,  $M_2$ , and  $M_3$ . The attributes of a product are production year (2020, 2021) and engine type (A, B). Time to failure between two activities is one month, where the  $M_1$ ,  $M_2$ , and  $M_3$  activities can occur in any month. Moreover, the first activity indicates the maintenance activity that occurred in the first month, and so on.

**Table 1.** Sample of collected data.

| Object No. | Production Year | Engine Type | First Activity | Second Activity | Third Activity |
|------------|-----------------|-------------|----------------|-----------------|----------------|
| 1          | 2020            | A           | $M_2$          | $M_1$           | $M_3$          |
| 2          | 2020            | B           | $M_1$          | $M_2$           | -              |
| 3          | 2021            | A           | $M_2$          | $M_3$           | $M_1$          |
| 4          | 2021            | A           | $M_2$          | $M_3$           | $M_1$          |

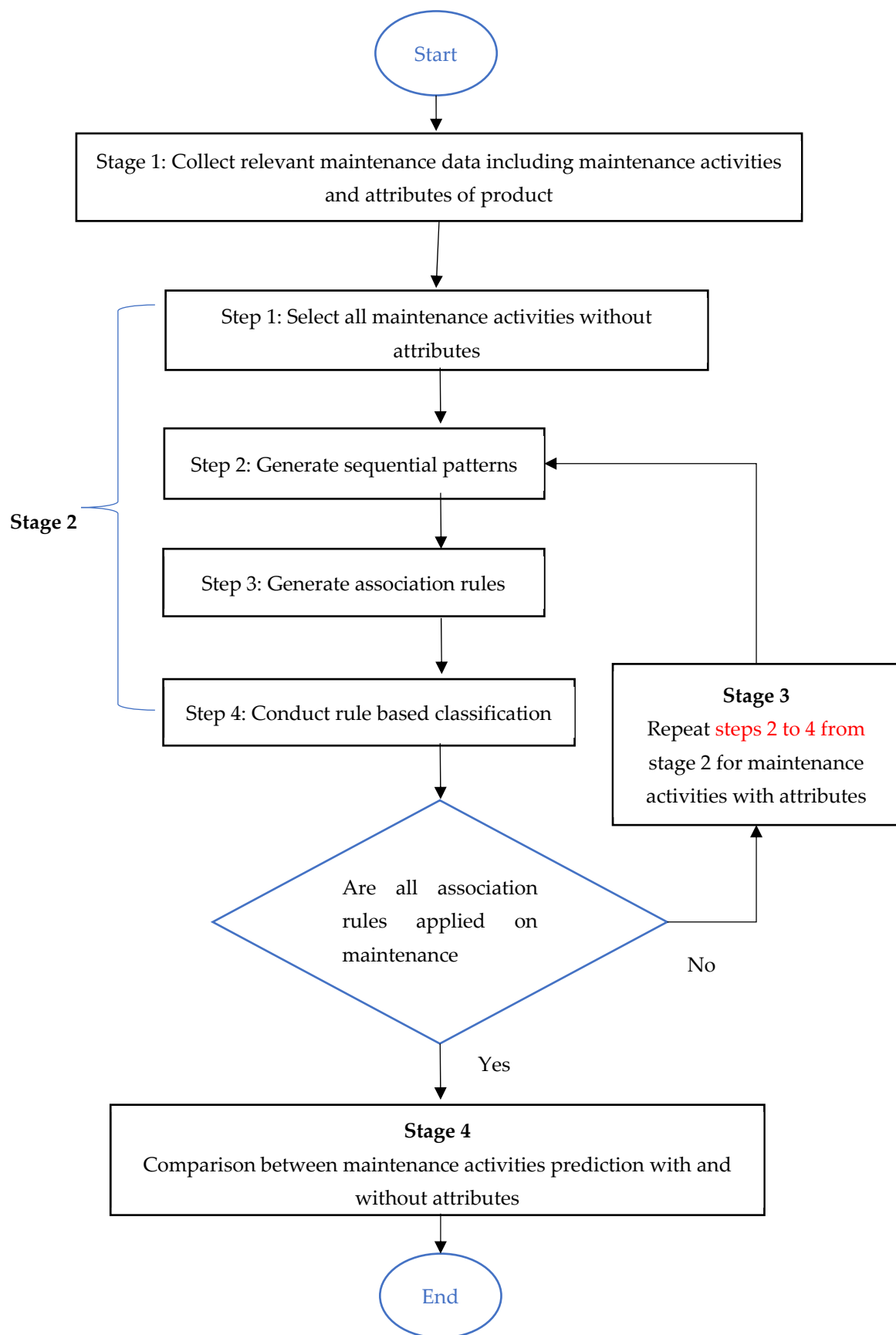


**Figure 1.** Research framework.

Stage 2. Maintenance activity prediction is conducted without attributes as follows:

Step 1. The types of maintenance activities which are applied on objects to maintain the product in a good performance are defined taking into consideration the time occurred of activity. From Table 1, three maintenance activities  $M_1$ ,  $M_2$ , and  $M_3$  will be analyzed.

Step 2. The Generalized Sequential Patterns (GSPs) are identified. A sequence database includes a sequence of ordered activities with or without time information. The sequence consists of multiple events where the event consists of one or more items in the transaction. For example,  $V = \{v_1, v_2, \dots, v_n\}$  is the set of activities,  $n$  is the number of activities, and the sequence is  $Q = (e_1, e_2, \dots, e_m)$ , in which  $e$  represent the events and  $m$  is the number of events in the transaction, given that  $e_2$  occurs after  $e_1$ . The possible number of patterns ( $N$ ) can be computed as follows [29]:

$$N = n^2 + \frac{n(n-1)}{2} \quad (1)$$

For illustration, three activities can occur in any maintenance activity number as shown in Table 2. So, ( $n = 9$ ) which indicates 117 possible patterns, can be produced as shown in Equation (1).

**Table 2.** Possible number of maintenance activities in series.

| First Activity | Second Activity | Third Activity |
|----------------|-----------------|----------------|
| $M_1$          | $M_1$           | $M_1$          |
| $M_2$          | $M_2$           | $M_2$          |
| $M_3$          | $M_3$           | $M_3$          |

Consequently, 117 possible patterns are obtained. Examples of possible patterns include the following:

(First activity =  $M_1$ , Second activity =  $M_2$ )

(First activity =  $M_3$ , Second activity =  $M_2$ )

(First activity =  $M_1$ , Second activity =  $M_2$ , Third activity =  $M_3$ )

Sequences are then classified as frequent or infrequent based on the support value (Sup.), which represents the percentage of activities related to the sequence rule of all objects as given in Equation (2). If the threshold of support value is too low, it leads to the involvement of too many items in patterns and rules, but if the support value is too high, it leads to fewer items involved in patterns and rules. So, the minimum support value is determined by the manufacturer based on the characteristics of the products [35]. When the support value of the sequence is more than or equal to the specified threshold value, the sequence is classified as a frequent sequential pattern. Figure 2 presents the GSP framework. The support value for events  $e_1$  and  $e_2$ ,  $\text{Sup}(e_1 \rightarrow e_2)$ , does not depend on the sequence and is calculated as follows:

$$\text{Sup}(e_1 \rightarrow e_2) = \frac{|e_1 \cup e_2|}{|R|} \times 100 \quad (2)$$

where  $e_1$  and  $e_2$  indicate the first event of sequence and the remaining events of sequence, respectively,  $|e_1 \cup e_2|$  is the number of objects that contains  $e_1$  and  $e_2$ , and  $|R|$  is the number of objects.

Note that

$$\text{Sup}(e_1 \rightarrow e_2) = \text{Sup}(e_2 \rightarrow e_1) \quad (3)$$

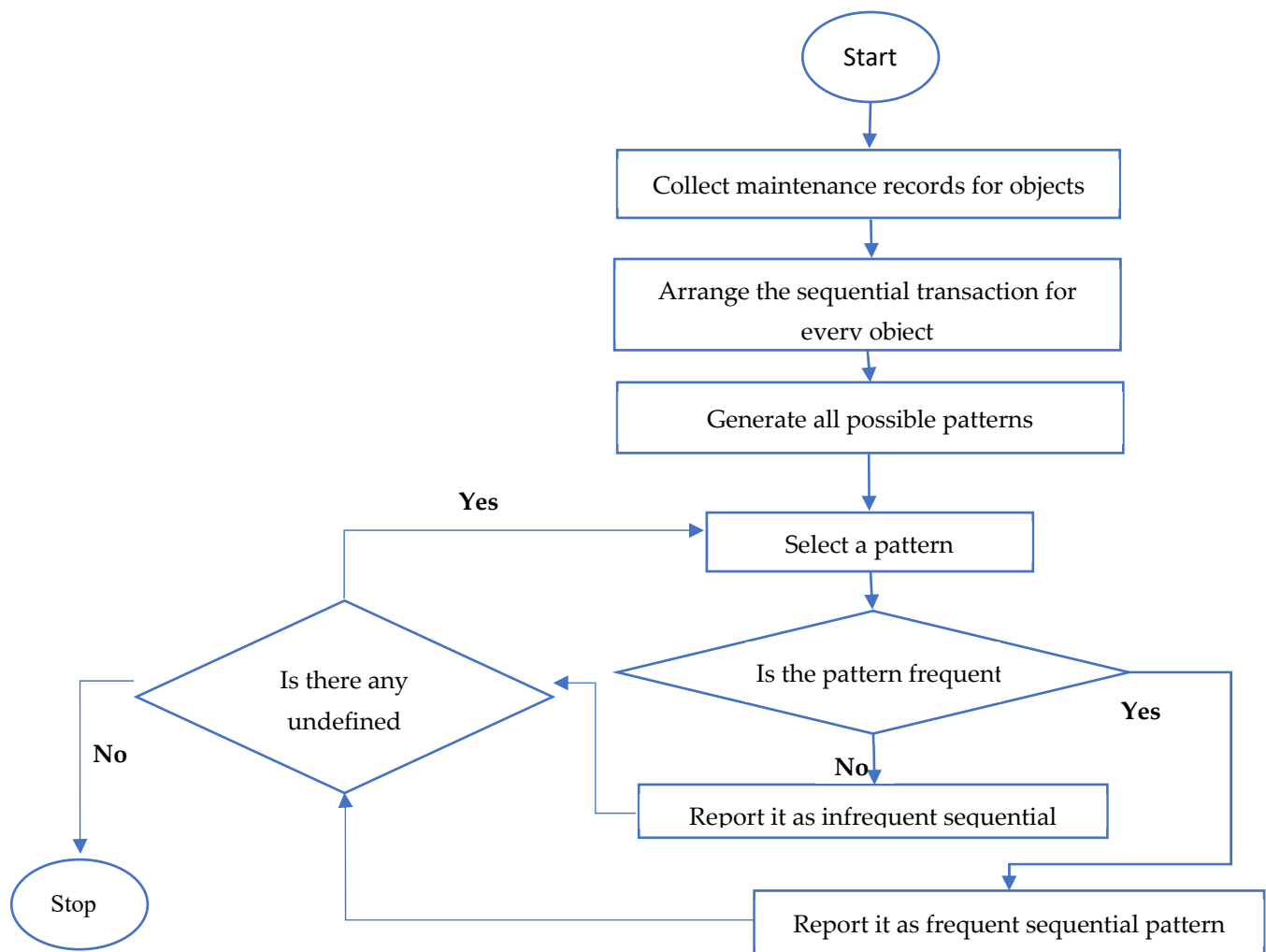


**Figure 2.** GSP framework.

The GSP for maintenance activities is used to identify the frequent patterns of maintenance activities. In Table 1, for example, one of the possible patterns is the pattern (First activity =  $M_2$ , Second activity =  $M_3$ ). The  $\text{Sup}(e_1 \rightarrow e_2)$  can then be computed using Equation (2) as follows:

$$\text{Sup}(e_1 \rightarrow e_2) = \frac{|e_1 \cup e_2|}{|R|} \times 100 = 50\%$$

where  $e_1$  and  $e_2$  refer to the first activity ( $M_2$ ) and the second activity ( $M_3$ ), respectively,  $|e_1 \cup e_2|$  is the number of objects ( $=2$ ) that contain the first activity of  $M_2$  and second activity of  $M_3$ , and  $|R|$  is the number of objects ( $=4$ ).

The  $\text{Sup}(e_1 \rightarrow e_2)$  of 50% indicates that 50% of all objects have the sequence pattern (First activity =  $M_2$ , Second activity =  $M_3$ ). All possible patterns are then defined and classified as frequent or infrequent patterns based on a predefined threshold support value. The best sequential patterns are further used to generate the rules association between items and finally determine the significant sequential patterns.

Step 3. The framework for generating the association rules is developed as shown in Figure 3. An association rule is a sequential relationship between the condition and decision activities. Condition activities (C) are the activities implemented for a specific object. Decision activities (D) are the activities on the same object that will be serviced at a later time. The Apriori algorithm is then employed for generating the association rules.

Typically, the association rule consists of IF (condition) and then (decision) statements. The condition and decision can be more than one activity. Statistical analysis including the support (Sup.) and confidence (Conf.) is applied to identify the significant association rules with the corresponding lift. The confidence ( $C \rightarrow D$ ) rule is the probability of the object having D given that it has C. The rule is significant when the confidence value is greater than or equal to a minimum confidence level determined by experts. It depends on the sequence of events so the Conf. ( $C \rightarrow D$ ) is different from Conf. ( $D \rightarrow C$ ). The Conf. value is computed using Equation (4).

$$\text{Conf. } (C \rightarrow D) = \frac{|C \cup D|}{|C|} \times 100\% \quad (4)$$

where  $|C \cup D|$  is the number of objects that contain C and D, while  $|C|$  is the number of objects that contain C. The lift or interest factor is calculated by the ratio of the joint probability of the events to the expected joint probability if they are statistically independent. A lift value less than or equal to one indicates that the rule is insignificant and the events are independent, while a lift value greater than one implies that the events are dependent. The lift ( $C \rightarrow D$ ) value is computed using Equation (5).

$$\text{Lift } (C \rightarrow D) = \frac{\text{Sup } (C \cup D)}{\text{Sup } (C) \cdot \text{Sup } (D)} \quad (5)$$

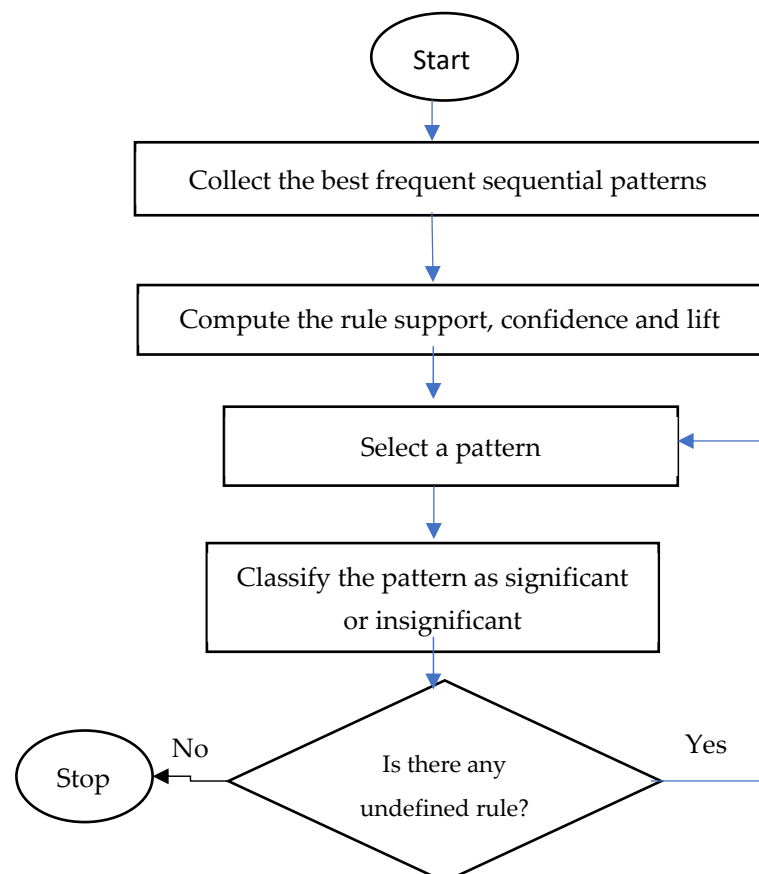


**Figure 3.** The framework of association rules significance testing.

The association rules can be generated for the maintenance activities to examine the sequential relationships between maintenance activities and determine the condition and decision activities. For illustration, in Table 1 the best frequent pattern is the (First activity =  $M_2$ , Second activity =  $M_3$ ) pattern. This pattern will be employed to associate

the rule. It has a support value of 50%. Given that, C indicates the first activity = M<sub>2</sub>, |C| is the number of objects that contain the first activity = M<sub>2</sub>, D is the second activity = M<sub>3</sub>, and |C U D| is the number of objects that contain the first and second activities, M<sub>2</sub> and M<sub>3</sub>, respectively. Then, the estimated values of the Conf. (C → D) and lift (C → D) are calculated using Equations (4) and (5), respectively, as follows:

$$\text{Conf. (C} \rightarrow \text{D)} = \frac{|C \cup D|}{|C|} \times 100\% = \frac{2}{3} \times 100\% = 66.7\%$$

$$\text{Sup (C U D)} = \text{Sup (First activity = M}_2\text{, Second activity = M}_3\text{)} = 50\%$$

$$\text{Sup (C)} = \text{Sup (First activity = M}_2\text{)} = 75\%$$

$$\text{Sup (D)} = \text{Sup (Second activity = M}_3\text{)} = 50\%$$

$$\text{Lift (C} \rightarrow \text{D)} = \frac{\text{Sup (C U D)}}{\text{Sup (C)} \cdot \text{Sup (D)}} = \frac{0.5}{0.75 \times 0.5} = 1.33$$

The Conf. (C → D) value indicates that 66.7% of the objects have the pattern (First activity = M<sub>2</sub>, and Second activity = M<sub>3</sub>). Suppose that the threshold Conf. is 60%, then this sequence is considered a significant pattern. The calculated lift value (=1.33) is greater than 1. Consequently, this rule is dependent. When a sequence consists of three events or more, many rules can be extracted from one rule by changing the condition and decision activities. All possible association rules should be classified into significant or insignificant rules. Only the significant rules are further employed to predict the next maintenance activities.

Step 4. Perform the rule-based classification of the maintenance activities using a collection of significant rules and evaluate the accuracy of prediction using coverage and accuracy measures. Let N<sub>v</sub> and N<sub>r</sub> denote the number of condition activities that satisfy the rule and the number of condition and decision activities that classify the rule, respectively. The |R| is the number of objects. Coverage is the percentage of the records that satisfies the condition of the rule as shown in Equation (6). Accuracy is the ratio of the number of condition and decision activities that classify the rule to the number of conditions as shown in Equation (7).

$$\text{Coverage} = \frac{N_v}{|R|} \times 100\% \quad (6)$$

$$\text{Accuracy} = \frac{N_r}{N_v} \times 100\% \quad (7)$$

For illustration, suppose that the testing data are collected for four objects as shown in Table 3. Suppose that the rule (First activity = M<sub>2</sub>); then, (Second activity = M<sub>3</sub>) is identified as a significant rule of maintenance activities.

**Table 3.** Testing data.

| Object No. | Production Year | Engine Type | First Activity | Second Activity | Third Activity |
|------------|-----------------|-------------|----------------|-----------------|----------------|
| 1          | 2020            | A           | M <sub>2</sub> | M <sub>3</sub>  | M <sub>3</sub> |
| 2          | 2020            | A           | M <sub>1</sub> | M <sub>2</sub>  | M <sub>3</sub> |
| 3          | 2020            | B           | M <sub>2</sub> | M <sub>3</sub>  | M <sub>1</sub> |
| 4          | 2021            | A           | M <sub>2</sub> | M <sub>3</sub>  | M <sub>1</sub> |

The  $N_v$  and  $N_r$  denote the number of objects that have (First activity =  $M_2$ ) that satisfied the rule and the number of objects that have (First activity =  $M_2$ , Second activity =  $M_3$ ) that classified the rule, respectively. Then, the coverage and accuracy are estimated as follows:

$$\text{Coverage} = \frac{N_v}{R} = \frac{3}{4} = 75\%$$

$$\text{Accuracy} = \frac{N_r}{N_v} = \frac{3}{3} = 100\%$$

The estimated coverage and accuracy values indicate that 75% of all objects have the first activity  $M_2$  and 100% of them validate the rule, respectively. All significant rules should be tested and evaluated.

Stage 3: Repeat steps (2–4) in stage 2 for maintenance activities with attributes. Each object has specific attributes that describe the properties of the object and define the physical or non-physical characteristics. In Table 1, the production year (2020, 2021) and engine type (A, B) are the attributes of objects. For illustration:

Step 1. The Generalize Sequential Pattern (GSP) is applied for planning the maintenance activities with attributes: production year only, engine type, or both attributes together. Then, all sequences should be analysed to identify the best sequential sequence for each attribute and both attributes. Table 4 represents the possible attributes of objects with either one attribute or both attributes.

**Table 4.** Possible attributes of objects with one attribute or both attributes.

| Attribute                               | Number of Attributes in One Sequence |
|-----------------------------------------|--------------------------------------|
| Production year = 2020                  | 1                                    |
| Production year = 2021                  | 1                                    |
| Engine type = A                         | 1                                    |
| Engine type = B                         | 1                                    |
| Production year = 2020, Engine type = A | 2                                    |
| Production year = 2020, Engine type = B | 2                                    |
| Production year = 2021, Engine type = A | 2                                    |
| Production year = 2021, Engine type = B | 2                                    |

One of the possible attributes that can be analysed with maintenance activities is engine type = A. Suppose that the best sequence of maintenance activities is (Engine type = A, First activity =  $M_2$ , Second activity =  $M_3$ ); then, the support value can be computed as follows:

$$\text{Sup}(e_1 \rightarrow e_2) = \frac{|e_1 \cup e_2|}{|R|} \times 100\% = \frac{2}{4} \times 100\% = 50\%,$$

where  $e_1$  indicates Engine type = A attribute,  $e_2$  indicates (First activity =  $M_2$ , Second activity =  $M_3$ ) sequence, and  $|e_1 \cup e_2|$  is the number of objects that contain Engine type = A, First activity =  $M_2$  and Second activity =  $M_3$ . The estimated  $\text{Sup}(e_1 \rightarrow e_2)$  value implies that 50% of all objects have the type of engine = A and a sequence (First activity =  $M_2$ , Second activity =  $M_3$ ).

Step 2. The association rules can be generated for the best sequence of maintenance activities with each possible attribute, that is, either one attribute or both. The condition is the attributes such as production year, engine type, or both, while the decision is the maintenance activities sequence. The rules are determined significant if the Conf. ( $C \rightarrow D$ ) value is more than or equal to a threshold value of Conf. Association rules on sequence (Engine type = A, First activity =  $M_2$ , Second activity =  $M_3$ ), which is considered

as best pattern for engine type = A. The Conf. ( $C \rightarrow D$ ) and lift values are calculated using Equations (4) and (5), respectively, as follows:

$$\text{Conf. } (C \rightarrow D) = \frac{|C \cup D|}{|C|} \times 100\% = \frac{2}{3} \times 100\% = 66.7\%$$

where C is Engine type = A attribute,  $|C|$  refers to the number of objects that contain Engine type = A, D indicates (First activity =  $M_2$ , Second activity =  $M_3$ ) sequence, and  $|C \cup D|$  refers to the objects that contain Engine type = A, First activity =  $M_2$ , Second activity =  $M_3$ .

$$\text{Lift } (C \rightarrow D) = \frac{\text{Sup } (C \cup D)}{\text{Sup } (C) \cdot \text{Sup } (D)} = \frac{0.5}{0.75 \times 0.5} = 1.33$$

where

$$\text{Sup } (C \cup D) = \text{Sup } (\text{Engine type} = A, \text{First activity} = M_2, \text{Second activity} = M_3) = 50\%$$

$$\text{Sup } (C) = \text{Sup } (\text{Engine type} = A) = 75\%$$

$$\text{Sup } (D) = \text{Sup } (\text{First activity} = M_2, \text{Second activity} = M_3) = 50\%$$

The calculated Conf. ( $C \rightarrow D$ ) means that 66.7% of objects that have Engine type = A, have (First activity =  $M_2$ , Second activity =  $M_3$ ) sequence. Suppose that the threshold of Conf. = 60%; then, this rule is significant with a lift value of more than 1, and thereby the rule is concluded as dependent.

Step 3. The rule-based classification is conducted by testing the significant rules of maintenance activities with each attribute using testing data in Table 3 and measuring the coverage and accuracy. One of the significant rules is an object which has Engine type = A, and the next maintenance activities sequence is (First activity =  $M_2$ , Second activity =  $M_3$ ). According to the testing data in Table 3, the coverage and accuracy values are calculated using Equations (6) and (7), respectively.

$$\text{Coverage} = \frac{N_v}{|R|} = \frac{3}{4} \times 100\% = 75\%$$

$$\text{Accuracy} = \frac{N_r}{N_v} = \frac{2}{3} \times 100\% = 66.7\%$$

where  $N_v$  is the number of (Engine type = A),  $N_r$  is the number of (First activity =  $M_2$ , Second activity =  $M_3$ ) that classified the rule, and  $|R|$  is the number of objects.

The estimated coverage indicates that 75% of objects are produced with Engine type = A and 66.7% of them validate the rule. All the significant rules should be tested and evaluated.

Stage 3. A comparison is made between the results in stages 1 and 2 after predicting maintenance activities with and without attributes.

### 3. Application on Maintenance Prediction

The proposed data mining framework will be illustrated using a real case study of washing machines produced by a single manufacturer. Washing machine products have many components such as tub, valves, pump and belt. These components require main maintenance activities to restore the operational functions of the component to their original states. Table 5 represents the components of the washing machine and maintenance activities.

The main attributes of the washing machine include (1) machine type {semi-automatic (SA), fully automatic (FA)} and (2) material of the drum {plastic, stainless steel}. Any combination of two attributes represents one type of washing machine. The data mining framework was applied to determine the significant frequent sequential patterns of maintenance activities with and without attributes and is presented as follows. In stage 1, the maintenance data were obtained from a washing machines manufacturer for 900 machines

for five months, including maintenance activities for tub, valves, pump, and belt components. The key attributes were the machine type and material of the drum as shown in Table 6. For example, machine No. 1 is a fully automatic machine with a plastic drum. The maintenance activity ( $M_3$ ) was performed on the belt in the first month then maintenance activity ( $M_2$ ) was applied on a pump in the second month. However, this machine did not require any maintenance activity in the third, fourth, and fifth months. The time ( $T$ ) between any two sequential activities is one month.

**Table 5.** Components of washing machine and maintenance activities.

| No. of Component | Component      | Maintenance Activity |
|------------------|----------------|----------------------|
| 1                | Tub and valves | $M_1$                |
| 2                | Pump           | $M_2$                |
| 3                | Belt           | $M_3$                |

**Table 6.** Collected data for washing machine.

| Machine No. | Machine Type | Material of Drum | Monthly Maintenance Activity |            |            |            |            |
|-------------|--------------|------------------|------------------------------|------------|------------|------------|------------|
|             |              |                  | Activity 1                   | Activity 2 | Activity 3 | Activity 4 | Activity 5 |
| 1           | FA           | Plastic          | $M_3$                        | $M_2$      |            |            |            |
| 2           | FA           | Stainless steel  | $M_3$                        | $M_2$      | $M_3$      | $M_3$      | $M_1$      |
| 3           | FA           | Plastic          | $M_3$                        | $M_2$      |            | $M_1$      | $M_3$      |
| 4           | FA           | Plastic          | $M_3$                        | $M_2$      | $M_3$      |            | $M_2$      |
| 5           | FA           | Stainless steel  | $M_1$                        | $M_3$      | $M_2$      | $M_1$      | $M_2$      |
| 6           | FA           | Stainless steel  | $M_2$                        |            | $M_3$      | $M_1$      | $M_2$      |
| 7           | FA           | Plastic          | $M_3$                        | $M_2$      | $M_3$      |            | $M_1$      |
| 8           | FA           | Plastic          | $M_3$                        | $M_2$      | $M_3$      | $M_2$      |            |
| 9           | FA           | Plastic          | $M_3$                        | $M_2$      |            | $M_1$      | $M_2$      |
| 10          | FA           | Plastic          | $M_3$                        | $M_2$      | $M_2$      | $M_3$      | $M_3$      |
| 11          | FA           | Plastic          | $M_3$                        | $M_2$      |            | $M_2$      | $M_1$      |
| 12          | FA           | Plastic          | $M_3$                        | $M_2$      |            | $M_1$      | $M_2$      |
| 13          | FA           | Stainless steel  | $M_1$                        | $M_3$      | $M_2$      | $M_2$      | $M_2$      |
| 14          | FA           | Stainless steel  | $M_3$                        | $M_2$      |            | $M_1$      | $M_2$      |
| ⋮           | ⋮            | ⋮                | ⋮                            | ⋮          | ⋮          | ⋮          | ⋮          |
| 885         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      |            | $M_3$      |
| 886         | SA           | Stainless steel  | $M_3$                        | $M_1$      | $M_2$      | $M_3$      | $M_1$      |
| 887         | SA           | Stainless steel  | $M_1$                        | $M_3$      | $M_2$      |            |            |
| 888         | SA           | Stainless steel  | $M_3$                        |            | $M_2$      | $M_3$      |            |
| 889         | SA           | Stainless steel  | $M_2$                        | $M_1$      | $M_2$      |            | $M_2$      |
| 890         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      |            | $M_3$      |
| 891         | SA           | Plastic          | $M_1$                        | $M_1$      | $M_2$      |            |            |
| 892         | SA           | Stainless steel  | $M_3$                        | $M_1$      | $M_2$      | $M_1$      | $M_3$      |
| 893         | SA           | Stainless steel  |                              | $M_1$      | $M_2$      | $M_2$      |            |
| 894         | SA           | Stainless steel  | $M_1$                        | $M_1$      | $M_3$      |            | $M_2$      |
| 895         | SA           | Stainless steel  | $M_1$                        |            | $M_2$      |            | $M_3$      |
| 896         | SA           | Stainless steel  |                              | $M_1$      |            |            |            |
| 897         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      | $M_1$      | $M_3$      |
| 898         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      |            |            |
| 899         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      |            |            |
| 900         | SA           | Plastic          | $M_1$                        | $M_3$      | $M_2$      | $M_3$      |            |

### 3.1. GSP for Maintenance Activities without Product Attributes

In Stage 2, the GSP analysis was conducted for the maintenance activities without attributes for five months. From Table 6, only monthly maintenance activities were analyzed. Then, the sequential patterns between maintenance activities were determined while considering the time occurrence of each activity. Each month refers to the maintenance activities  $M_1$ ,  $M_2$ , and  $M_3$  for tub and valves, pump, and belt components, respectively.

Hence, 3 activities could occur in each month, and thereby the number of activities was 15 activities for five months. Table 7 represents all possible activities for five months. For example, Activity 1 =  $M_1$  is one of the possible activities.

**Table 7.** All possible activities for five months.

| Activity 1 | Activity 2 | Activity 3 | Activity 4 | Activity 5 |
|------------|------------|------------|------------|------------|
| $M_1$      | $M_1$      | $M_1$      | $M_1$      | $M_1$      |
| $M_2$      | $M_2$      | $M_2$      | $M_2$      | $M_2$      |
| $M_3$      | $M_3$      | $M_3$      | $M_3$      | $M_3$      |

Then, the number of possible patterns (=330) considering the number of activities ( $n = 15$ ) was calculated using Equation (1). Each pattern has a support value that represents the frequency of the pattern according to all machines. Each pattern has a support value representing the frequency of the pattern according to all machines. The threshold of the Sup. Value is imposed based on product attributes and the manufacturer's opinion [35]. Further, the pattern is more significant when the Sup. value increases [34]. Hence, the support threshold was determined based on the manufacturer's opinion and set as 20% to enhance the accuracy. That is, a sequential pattern that had a support value of 20% or more is identified as a frequent sequential pattern. RapidMiner software was used to generate all sequential patterns. Table 8 lists the obtained frequent sequential patterns ranked in descending order based on their estimated support values (Sup. (100%)).

**Table 8.** Frequent sequential patterns of maintenance activities.

| No. of Frequent Pattern | Frequent Sequential Pattern                                        | No. of Items Per Sequence | No. of Repeated Pattern | Sup. (100%) |
|-------------------------|--------------------------------------------------------------------|---------------------------|-------------------------|-------------|
| 1                       | Activity 1 = $M_1$ ,<br>Activity 2 = $M_3$                         | 2                         | 303                     | 33.7        |
| 2                       | Activity 2 = $M_1$ ,<br>Activity 3 = $M_2$                         | 2                         | 300                     | 33.3        |
| 3                       | Activity 1 = $M_1$ ,<br>Activity 3 = $M_2$                         | 2                         | 267                     | 29.7        |
| 4                       | Activity 2 = $M_3$ ,<br>Activity 3 = $M_2$                         | 2                         | 254                     | 28.2        |
| 5                       | Activity 1 = $M_1$ ,<br>Activity 2 = $M_3$ ,<br>Activity 3 = $M_2$ | 3                         | 251                     | 27.9        |
| 6                       | Activity 3 = $M_2$ ,<br>Activity 4 = $M_1$                         | 2                         | 249                     | 27.7        |
| 7                       | Activity 4 = $M_1$ ,<br>Activity 5 = $M_2$                         | 2                         | 213                     | 23.7        |
| 8                       | Activity 1 = $M_3$ ,<br>Activity 2 = $M_2$                         | 2                         | 191                     | 21.2        |

From Table 8, the largest support (=33.7%) corresponds to the two-item sequence (Activity 1 =  $M_1$ , Activity 2 =  $M_3$ ). That is, about 303 out of 900 machines have the same sequence of maintenance activities (Activity 1 =  $M_1$ , Activity 2 =  $M_3$ ). The best three-item sequence is the sequence (Activity 1 =  $M_1$ , Activity 2 =  $M_3$ , Activity 3 =  $M_2$ ), which has a support value of 27.9%. Next, the IF-THEN rules were generated for the best patterns to identify the relationships between activities. The confidence threshold value was determined based on the manufacturer and expert opinion, taking into consideration the characteristics of the product [20]. Consequently, the rules were judged significant when the confidence and lift values are more than 80% and more than one, respectively. Table 9 lists the association rules of the best pattern for two and three items.

**Table 9.** Association rules according to the best patterns.

| No. of Rule | Condition (C)                                             | Decision (D)                                                 | Sup. 100% | Conf. 100% | Lift | Significance  |
|-------------|-----------------------------------------------------------|--------------------------------------------------------------|-----------|------------|------|---------------|
| 1           | Activity 1 = M <sub>1</sub>                               | Activity 2 = M <sub>3</sub>                                  | 33.7      | 92.9       | 2.72 | Significant   |
| 2           | Activity 1 = M <sub>1</sub>                               | Activity 2 = M <sub>3</sub> ,<br>Activity 3 = M <sub>2</sub> | 27.9      | 77.0       | 2.73 | Insignificant |
| 3           | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> | Activity 3 = M <sub>2</sub>                                  | 27.9      | 82.8       | 1.28 | Significant   |

From Table 9, the first rule has the condition (Activity 1 = M<sub>1</sub>), decision (Activity 2 = M<sub>3</sub>), support value = 33.7%, Conf. = 92.9%, and a lift of 2.72. This rule is significant because the confidence value is more than 80%. Similarly, rule 3 is considered a significant rule. Both significant rules can be used to predict the next maintenance activity. That is, if M<sub>1</sub> activity occurred in the first month (Activity 1 = M<sub>1</sub>), then the next maintenance activity in the second month will be M<sub>3</sub>. If the sequence (Activity 1 = M<sub>1</sub>, Activity 2 = M<sub>3</sub>) takes place in the first and second months, then maintenance activity M<sub>2</sub> will be performed in the third month. Finally, testing data were collected for 400 machines with related maintenance activities and attributes to evaluate the accuracy of prediction for significant rules as shown in Table 10. The attributes were machine type (SA, FA) and drum material (plastic and stainless steel). The testing data were classified using coverage and accuracy measures. The rules were accurate when coverage and accuracy are more than 20% and 80%, respectively.

**Table 10.** Testing data for washing machine.

| Machine No. | Machine Type | Material of Drum | Monthly Maintenance Activity |                |                |                |                |
|-------------|--------------|------------------|------------------------------|----------------|----------------|----------------|----------------|
|             |              |                  | Activity 1                   | Activity 2     | Activity 3     | Activity 4     | Activity 5     |
| 1           | FA           | Stainless steel  | M <sub>3</sub>               | M <sub>2</sub> | M <sub>3</sub> | M <sub>3</sub> | M <sub>1</sub> |
| 2           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> |                | M <sub>1</sub> | M <sub>3</sub> |
| 3           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> | M <sub>3</sub> |                | M <sub>2</sub> |
| 4           | FA           | Stainless steel  | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> | M <sub>1</sub> | M <sub>2</sub> |
| 5           | FA           | Stainless steel  | M <sub>2</sub>               |                | M <sub>3</sub> | M <sub>1</sub> | M <sub>2</sub> |
| 6           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> | M <sub>3</sub> |                | M <sub>1</sub> |
| 7           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> | M <sub>3</sub> | M <sub>2</sub> |                |
| 8           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> |                | M <sub>1</sub> | M <sub>2</sub> |
| 9           | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> | M <sub>2</sub> | M <sub>3</sub> | M <sub>3</sub> |
| 10          | FA           | Plastic          | M <sub>3</sub>               | M <sub>2</sub> | M <sub>3</sub> |                | M <sub>1</sub> |
| 11          | FA           | Stainless steel  | M <sub>3</sub>               |                |                | M <sub>1</sub> |                |
| 12          | FA           | Stainless steel  | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> | M <sub>1</sub> | M <sub>2</sub> |
| 13          | FA           | Stainless steel  |                              | M <sub>1</sub> | M <sub>2</sub> | M <sub>1</sub> | M <sub>2</sub> |
| 14          | FA           | Stainless steel  | M <sub>3</sub>               |                |                | M <sub>1</sub> | M <sub>3</sub> |
| ⋮           | ⋮            | ⋮                | ⋮                            | ⋮              | ⋮              | ⋮              | ⋮              |
| 393         | SA           | Plastic          | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> | M <sub>2</sub> | M <sub>1</sub> |
| 394         | SA           | Plastic          | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> | M <sub>2</sub> |                |
| 395         | SA           | Plastic          | M <sub>1</sub>               |                | M <sub>2</sub> |                | M <sub>3</sub> |
| 396         | SA           | Plastic          | M <sub>1</sub>               | M <sub>3</sub> |                |                |                |
| 397         | SA           | Stainless steel  | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> |                | M <sub>2</sub> |
| 398         | SA           | Plastic          | M <sub>1</sub>               | M <sub>3</sub> | M <sub>2</sub> |                | M <sub>3</sub> |
| 399         | SA           | Stainless steel  | M <sub>3</sub>               | M <sub>1</sub> |                |                | M <sub>1</sub> |
| 400         | SA           | Stainless steel  |                              | M <sub>3</sub> | M <sub>2</sub> |                | M <sub>1</sub> |

Table 11 represents the prediction results for the testing data. For the first rule (Activity 1 = M<sub>1</sub> then Activity 2 = M<sub>3</sub>), Activity 1 = M<sub>1</sub> was repeated for 150 machines, in which the sequence Activity 1 = M<sub>1</sub>; then, Activity 2 = M<sub>3</sub> was repeated for 135 machines. The estimated coverage and accuracy are 37.5% and 90.0%. That is, Activity 1 = M<sub>1</sub> was covered for 37.5% of machines of which 91.9% have the next Activity 2 = M<sub>3</sub>. These results indicate that the two rules were predicted accurately based on rule coverage and accuracy.

**Table 11.** Prediction results of testing data.

| No. | Condition (C)                                                | Decision (D)                | N <sub>v</sub> | N <sub>r</sub> | Coverage % | Accuracy % |
|-----|--------------------------------------------------------------|-----------------------------|----------------|----------------|------------|------------|
| 1   | Activity 1 = M <sub>1</sub>                                  | Activity 2 = M <sub>3</sub> | 150            | 135            | 37.5       | 90.0       |
| 2   | Activity 1 = M <sub>1</sub> ,<br>Activity 2 = M <sub>3</sub> | Activity 3 = M <sub>2</sub> | 135            | 124            | 33.8       | 91.9       |

### 3.2. GSP for Maintenance Activities with Product Attributes

In Stage 3, the GSP analysis was conducted for maintenance activities with washing machine attributes including machine-type attributes (fully-automatic (FA) or semi-automatic (SA)), and the material of the drum (plastic or stainless steel). Table 12 displays the frequent sequential patterns with the machine-type attribute. The best sequential pattern is the frequent sequence that corresponds to the largest number of items and the highest support value. Consequently, the sequence (Machine type = SA, Activity 1 = M<sub>1</sub>, Activity 2 = M<sub>3</sub>, Activity 3 = M<sub>2</sub>) is found the best four-item sequence (support value = 20.6%) for the SA machine, whereas the sequence (Machine type = FA, Activity 4 = M<sub>1</sub>, Activity 5 = M<sub>2</sub>) is found the best three-item sequence (support value of 23.7%) for the FA machines.

**Table 12.** Frequent sequential patterns for machine types.

| No. | Sequential Pattern                                                                             | Sequence Items Per | Repeated Patterns | Sup. (100%) |
|-----|------------------------------------------------------------------------------------------------|--------------------|-------------------|-------------|
| 1   | SA, Activity 3 = M <sub>2</sub>                                                                | 2                  | 469               | 52.1        |
| 2   | SA, Activity 2 = M <sub>1</sub>                                                                | 2                  | 291               | 32.3        |
| 3   | SA, Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                                  | 3                  | 275               | 30.6        |
| 4   | SA, Activity 1 = M <sub>1</sub>                                                                | 2                  | 203               | 22.6        |
| 5   | SA, Activity 2 = M <sub>3</sub>                                                                | 2                  | 195               | 21.7        |
| 6   | SA, Activity 1 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                                  | 3                  | 194               | 21.6        |
| 7   | SA, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub>                                  | 3                  | 192               | 21.3        |
| 8   | SA, Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub>                                  | 3                  | 187               | 20.8        |
| 9   | SA, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> ,<br>Activity 3 = M <sub>2</sub> | 4                  | 185               | 20.6        |
| 1   | FA, Activity 5 = M <sub>2</sub>                                                                | 2                  | 289               | 32.1        |
| 2   | FA, Activity 4 = M <sub>1</sub>                                                                | 2                  | 226               | 25.1        |
| 3   | FA, Activity 1 = M <sub>3</sub>                                                                | 2                  | 219               | 24.3        |
| 4   | FA, Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub>                                  | 3                  | 213               | 23.7        |
| 5   | FA, Activity 2 = M <sub>2</sub>                                                                | 2                  | 194               | 21.6        |
| 6   | FA, Activity 1 = M <sub>3</sub> , Activity 2 = M <sub>2</sub>                                  | 3                  | 190               | 21.1        |

Further, the sequential patterns for the material of the drum are generated as shown in Table 13. It is noted that the best four-item sequential pattern for plastic material is (Material of drum = plastic, Activity 1 = M<sub>1</sub>, Activity 2 = M<sub>3</sub>, Activity 3 = M<sub>2</sub>) with a support value of 20.2%. However, for the stainless steel attribute, the best three-item sequence is (Material of drum = stainless steel, Activity 2 = M<sub>1</sub>, Activity 3 = M<sub>2</sub>) with a support value of 33.1%.

Finally, Table 14 displays the frequent sequential patterns for both machine-type = SA and material of drum attributes. It is found that the five-item sequence (Machine type = SA, Material of drum = plastic, Activity 1 = M<sub>1</sub>, Activity 2 = M<sub>3</sub>, Activity 3 = M<sub>2</sub>) is the best sequence, with a support value of 20.2%. For machine type = SA and material of drum = stainless steel attribute, however, the four-item sequence (Machine type = SA, Material of drum = stainless steel, Activity 2 = M<sub>1</sub>, Activity 3 = M<sub>2</sub>) is the best sequence, with a support value of 30.4%.

**Table 13.** Frequent sequential patterns for attributes of drum material.

| No. | Sequential Pattern                                                                               | Sequence Items Per | Repeated Patterns | Sup. (100%) |
|-----|--------------------------------------------------------------------------------------------------|--------------------|-------------------|-------------|
| 1   | plastic, Activity 3 = M <sub>2</sub>                                                             | 2                  | 197               | 21.9        |
| 2   | plastic, Activity 1 = M <sub>1</sub>                                                             | 2                  | 192               | 21.3        |
| 3   | plastic, Activity 2 = M <sub>3</sub>                                                             | 2                  | 191               | 21.2        |
| 4   | plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub>                               | 3                  | 189               | 21.0        |
| 5   | plastic, Activity 1 = M <sub>3</sub>                                                             | 2                  | 185               | 20.6        |
| 6   | plastic, Activity 2 = M <sub>2</sub>                                                             | 2                  | 185               | 20.6        |
| 7   | plastic, Activity 1 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                               | 3                  | 184               | 20.4        |
| 8   | plastic, Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub>                               | 3                  | 183               | 20.3        |
| 9   | plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 4                  | 182               | 20.2        |
| 10  | plastic, Activity 1 = M <sub>3</sub> , Activity 2 = M <sub>2</sub>                               | 3                  | 181               | 20.1        |
| 1   | stainless steel, Activity 3 = M <sub>2</sub>                                                     | 2                  | 386               | 42.9        |
| 2   | stainless steel, Activity 4 = M <sub>1</sub>                                                     | 2                  | 339               | 37.7        |
| 3   | stainless steel, Activity 2 = M <sub>1</sub>                                                     | 2                  | 315               | 35.0        |
| 4   | stainless steel, Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                       | 3                  | 298               | 33.1        |
| 5   | stainless steel, Activity 5 = M <sub>2</sub>                                                     | 2                  | 209               | 23.2        |
| 6   | stainless steel, Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub>                       | 3                  | 198               | 22.0        |
| 7   | stainless steel, Activity 3 = M <sub>2</sub> , Activity 4 = M <sub>1</sub>                       | 3                  | 226               | 25.1        |

**Table 14.** Frequent sequential patterns for machine type = SA and the material of drum.

| No. | Sequential Pattern                                                                                   | Sequence Items | Repeated Pattern | Sup. (100%) |
|-----|------------------------------------------------------------------------------------------------------|----------------|------------------|-------------|
| 1   | SA, plastic, Activity 1 = M <sub>1</sub>                                                             | 3              | 191              | 21.2        |
| 2   | SA, plastic, Activity 2 = M <sub>3</sub>                                                             | 3              | 190              | 21.1        |
| 3   | SA, plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub>                               | 4              | 189              | 21.0        |
| 4   | SA, plastic, Activity 3 = M <sub>2</sub>                                                             | 3              | 184              | 20.4        |
| 5   | SA, plastic, Activity 1 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                               | 4              | 184              | 20.4        |
| 6   | SA, plastic, Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub>                               | 4              | 182              | 20.2        |
| 7   | SA, plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 5              | 182              | 20.2        |
| 1   | SA, stainless steel, Activity 2 = M <sub>1</sub>                                                     | 3              | 290              | 32.2        |
| 2   | SA, stainless steel, Activity 3 = M <sub>2</sub>                                                     | 3              | 285              | 31.7        |
| 3   | SA, stainless steel, Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                       | 4              | 274              | 30.4        |

Table 15 shows the frequent sequential patterns for machine-type = FA and the attributes of the drum material. It is found that the sequence (Machine-type = FA, Material of drum = plastic, Activity 1 = M<sub>3</sub>, Activity 2 = M<sub>2</sub>) is the best four-item sequence with a support value of 20.1%, while for machine type = FA and material of drum = stainless steel attribute, the best four-item sequence is (Machine-type = FA, Material of drum = stainless steel, Activity 4 = M<sub>1</sub>, Activity 5 = M<sub>2</sub>) with a support value of 22.0%.

Next, the IF-THEN rules were generated to identify the relationships between maintenance activities and attributes for patterns that were selected for maintenance activities and attributes as shown in Table 16. It is noted that only the sequences (SA, plastic Activity 1 = M<sub>1</sub>, Activity 2 = M<sub>3</sub>, Activity 3 = M<sub>2</sub>), (SA, stainless steel, Activity 2 = M<sub>1</sub>, Activity 3 = M<sub>2</sub>), and (FA, stainless steel, Activity 4 = M<sub>1</sub>, Activity 5 = M<sub>2</sub>) are the significant rules.

**Table 15.** Frequent sequential patterns for machine type = FA and material of drum.

| No. | Sequential Pattern                                                             | Sequence Items | Repeated Patterns | Sup. (100%) |
|-----|--------------------------------------------------------------------------------|----------------|-------------------|-------------|
| 1   | FA, plastic, Activity 2 = M <sub>2</sub>                                       | 3              | 185               | 20.6        |
| 2   | FA, plastic, Activity 1 = M <sub>3</sub>                                       | 3              | 185               | 20.6        |
| 3   | FA, plastic, Activity 1 = M <sub>3</sub> , Activity 2 = M <sub>2</sub>         | 4              | 181               | 20.1        |
| 1   | FA, stainless steel, Activity 4 = M <sub>1</sub>                               | 3              | 208               | 23.1        |
| 2   | FA, stainless steel, Activity 5 = M <sub>2</sub>                               | 3              | 200               | 22.2        |
| 3   | FA, stainless steel, Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub> | 4              | 198               | 22.0        |

**Table 16.** Association rules for best patterns of each possible attribute.

| Rule | Condition (C)       | Decision (D)                                                                            | Sup. 100% | Conf. 100% | Lift | Significance  |
|------|---------------------|-----------------------------------------------------------------------------------------|-----------|------------|------|---------------|
| 1    | SA                  | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 20.6      | 37.5       | 1.35 | Insignificant |
| 2    | FA                  | Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub>                               | 23.7      | 52.3       | 2.21 | Insignificant |
| 3    | plastic             | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 20.2      | 47.9       | 1.72 | Insignificant |
| 4    | SA, plastic         | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 20.2      | 94.8       | 3.40 | Significant   |
| 5    | stainless steel     | Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                               | 33.1      | 57.3       | 1.91 | Insignificant |
| 6    | SA, stainless steel | Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                               | 30.4      | 91.0       | 2.73 | Significant   |
| 7    | FA, plastic         | Activity 1 = M <sub>3</sub> , Activity 2 = M <sub>2</sub>                               | 20.1      | 47.6       | 2.24 | Insignificant |
| 8    | FA, stainless steel | Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub>                               | 22.0      | 90.4       | 3.82 | Significant   |

In addition to the three significant rules in Table 16, it is found that when machine-type is SA, there is a 95.1% probability of M<sub>2</sub> occurring in the third month. In addition, the sequence (Machine type = SA, Activity 3 = M<sub>2</sub>) was repeated for 469 machines from 900 machines, thus the Sup. value is 52.1%. This sequence is consequently included in the list of significant sequences shown in Table 17. Measuring coverage and accuracy of the rules, the rules are found to be accurate.

**Table 17.** Prediction results of significant rules of maintenance activities with attributes.

| No. | Condition (C)       | Decision (D)                                                                            | N <sub>v</sub> | N <sub>r</sub> | Coverage % | Accuracy % |
|-----|---------------------|-----------------------------------------------------------------------------------------|----------------|----------------|------------|------------|
| 1   | SA                  | Activity 3 = M <sub>2</sub>                                                             | 250            | 227            | 62.5       | 90.8       |
| 2   | SA, plastic         | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> | 140            | 130            | 35.0       | 92.9       |
| 3   | SA, stainless steel | Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub>                               | 110            | 99             | 27.5       | 90.0       |
| 4   | FA, stainless steel | Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub>                               | 120            | 109            | 30.0       | 90.8       |

Stage 3. A comparison is made between the results of maintenance activities with and without attributes as shown in Table 18. Generally, the prediction of maintenance activities without product attributes provides a general and less accurate sequence that helps in the prediction of the next activity. However, examining the product attributes aims at finding the major causes by focusing on the major attributes in maintenance prediction and hence provides a high-accuracy sequence.

**Table 18.** Comparison of maintenance activities analysis with and without attributes.

| No. | Aspect        | Maintenance Activities Analysis without Attributes                                                                                                                                                     | Maintenance Activities Analysis with Attributes                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | Best sequence | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub>                                                                                                                | Depending on possible attributes, either one attribute or both<br>(Machine-type = SA, Activity 3 = M <sub>2</sub> ),<br>(Machine type = SA and Material of drum = plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub> ),<br>(Machine type = SA and Material of drum = stainless steel, Activity 2 = M <sub>1</sub> , Activity 3 = M <sub>2</sub> ),<br>(Machine type = FA and Material of drum = stainless steel, Activity 4 = M <sub>1</sub> , Activity 5 = M <sub>2</sub> ) |
| 2   | Accuracy      | Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub><br>Accuracy = 91.9                                                               | Machine type = SA and Material of drum = plastic, Activity 1 = M <sub>1</sub> , Activity 2 = M <sub>3</sub> , Activity 3 = M <sub>2</sub><br>Accuracy = 92.9                                                                                                                                                                                                                                                                                                                                                               |
| 3   | Importance    | Helps in preparation and prediction of next activity when the first activity in the first month occurred or when the first activity in the first month and second activity in the third month occurred | When one attribute is considered, (Machine type = SA) is the major attribute, and when both attributes are considered, (Machine type = SA and Material of drum = plastic, Machine type = SA and Material of drum = stainless steel, Machine type = FA and Material of drum = stainless steel) were the major product attributes                                                                                                                                                                                            |

#### 4. Research Results

According to the results concluded from Table 18, the best three-item sequence of maintenance activities without attributes indicates that the tub and valve repair occurs in the first month, belt repair occurs in the second month, and pump repair occurs in the third month. This rule helps in predicting the sequence of maintenance activities and the prediction of the next maintenance activity. On the other hand, analysing maintenance activities with attributes helps in determining the critical product attributes. For this case study, the maintenance manager should focus on semi-automatic machines that, in general, need pump repair in the third month. A semi-automatic machine with a plastic drum requires tub and valve repair in the first month, belt repair in the second month, and pump repair in the third month. A semi-automatic machine with a stainless steel drum requires tub and valves repair in the second month, and pump repair in the third month, while the fully automatic machine with a stainless steel drum needs tub and valves repair in the fourth month and pump repair in the fifth month. Considering the product attributes, root cause analysis should be performed for each major attribute, thereby reducing downtime and maintenance costs.

Based on research findings, the proposed framework offers the following advantages: (i) prediction of the frequent sequential patterns of maintenance activities without attributes; (ii) considering the time of maintenance activity; (iii) generating association rules between activities to determine the importance of rules; (iv) conducting rule-based classification to evaluate the accuracy of the rules; (v) analysing maintenance activities with attributes to examine their effects and predict maintenance activities based on attributes; (vi) generates association rules between activities and attributes; and (vii) performing rule-based classification to evaluate the accuracy of the rules with attributes. These advantages may encourage maintenance engineering in adopting this framework for predicting maintenance activities for a wide range of applications with and without product attributes.

#### 5. Conclusions

Many efforts have been made to predict the sequence of maintenance activities based on product attributes and develop effective preventive maintenance plans. The research, therefore, proposed a data mining framework to predict the sequential pattern of maintenance activities.

nance activities with and without attributes, taking into consideration occurred time by using GSP and associating rules between maintenance activities. A case study of washing machines was conducted for illustration. The results showed that the proposed framework is effective in predicting the sequence of maintenance activities with and without machine attributes and may provide valuable support to product and maintenance engineers in predicting the maintenance sequence and the next maintenance activities and developing an effective maintenance plan. A comparison between the results of the prediction results with and without product attributes was made. It is concluded that the inclusion of product attributes provides more accurate sequences and valuable information regarding the root causes of maintenance problems. In addition, the proposed framework can support industrial firms in reducing the probability and impacts of failures, thereby reducing economic losses due to maintenance activities. Future research considers extending the framework for predicting needed spare parts and manpower.

**Author Contributions:** Conceptualization, methodology, A.A.-R. and B.A.H.; validation, A.A.-R., B.A.H. and N.L.; formal analysis, A.A.-R., B.A.H. and N.L.; investigation, A.A.-R. and B.A.H.; resources, A.A.-R., B.A.H. and N.L.; data curation, A.A.-R., B.A.H. and N.L.; writing—original draft preparation, A.A.-R., B.A.H. and N.L.; writing—review and editing, A.A.-R., B.A.H. and N.L.; visualization, A.A.-R. and B.A.H.; supervision, A.A.-R., B.A.H. and N.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Huang, J.; Chang, Q.; Zou, J.; Arinez, J. A Real-Time Maintenance Policy for Multi-Stage Manufacturing Systems Considering Imperfect Maintenance Effects. *IEEE Access* **2018**, *6*, 62174–62183. [\[CrossRef\]](#)
- Al-Refaie, A.; Lepkova, N.; Camlibel, M.E. The Relationships between the Pillars of TPM and TQM and Manufacturing Performance Using Structural Equation Modeling. *Sustainability* **2022**, *14*, 1497. [\[CrossRef\]](#)
- Al-Refaie, A.; Al-Hawadi, A. Optimal fuzzy repairs' scheduling and sequencing of failure types over multiple periods. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 201–217. [\[CrossRef\]](#)
- Halmetoja, E.; Lepkova, N. Utilising Building Information Models in Facility Maintenance and Operations. *Tek. Dergi* **2022**, *33*, 12351–12337. [\[CrossRef\]](#)
- Al-Refaie, A.; Almowas, H. Multi-objective maintenance planning under preventive maintenance. *J. Qual. Maint. Eng.* **2023**, *1*, 50–70. [\[CrossRef\]](#)
- Sun, Y.; Sun, Z. Selective Maintenance on a Multi-State Transportation System Considering Maintenance Sequence Arrangement. *IEEE Access* **2021**, *9*, 70048–70060. [\[CrossRef\]](#)
- Al-Refaie, A.; Al-Shalalkeh, H.; Lepkova, N. Proposed procedure for optimal maintenance scheduling under emergent failures. *J. Civ. Eng. Manag.* **2020**, *26*, 396–409. [\[CrossRef\]](#)
- Yu, T.; Zhu, C.; Chang, Q.; Wang, J. Imperfect corrective maintenance scheduling for energy efficient manufacturing systems through online task allocation method. *J. Manuf. Syst.* **2019**, *53*, 282–290. [\[CrossRef\]](#)
- Junqueira, V.S.V.; Nagano, M.S.; Miyata, H.H. Procedure structuring for programming aircraft maintenance activities. *Revista Gestão* **2020**, *27*, 2–20. [\[CrossRef\]](#)
- Kaparthi, S.; Bumblauskas, D. Designing predictive maintenance systems using decision tree-based machine learning techniques. *Int. J. Qual. Reliab. Manag.* **2020**, *37*, 659–686. [\[CrossRef\]](#)
- Kumar, G.; Jain, V.; Gandhi, O.P. Steady-state availability analysis of repairable mechanical systems with opportunistic maintenance by using Semi-Markov process. *Int. J. Syst. Assur. Eng. Manag.* **2014**, *5*, 664–678. [\[CrossRef\]](#)
- Patel, J.K. The Importance of Equipment Maintenance Forecasting. *Int. J. Mech. Eng.* **2021**, *8*, 7–11. [\[CrossRef\]](#)
- Mosallam, A.; Byttner, S.; Svensson, M.; Rognvaldsson, T. Nonlinear Relation Mining for Maintenance Prediction. In Proceedings of the 2011 Aerospace Conference, Big Sky, MT, USA, 5–12 March 2011; pp. 1–9. [\[CrossRef\]](#)
- Wang, H.; Gao, J.; Wu, H. Direct maintenance cost prediction of civil aircraft. *Aircr. Eng. Aerosp. Technol. Int. J.* **2014**, *86*, 406–414. [\[CrossRef\]](#)

15. Lee, H.; Kim, C.O.; Ko, H.H.; Kim, M.-K. Yield Prediction Through the Event Sequence Analysis of the Die Attach Process. *IEEE Trans. Semicond. Manuf.* **2015**, *28*, 563–570. [\[CrossRef\]](#)
16. Jacobs, J.A.; Mathews, M.J.; Kleingeld, M. Failure Prediction of Mine Compressors. *J. Fail. Anal. Prev.* **2019**, *19*, 976–985. [\[CrossRef\]](#)
17. Wen, Q.; Zhang, J.-P.; Hu, Z.-Z.; Xiang, X.-S.; Shi, T. A Data-Driven Approach to Improve the Operation and Maintenance Management of Large Public Buildings. *IEEE Access* **2019**, *7*, 176127–176140. [\[CrossRef\]](#)
18. Wakiru, J.; Pintelon, L.; Muchiri, P.; Chemweno, P. A data mining approach for lubricant-based fault diagnosis. *J. Qual. Maint. Eng.* **2021**, *27*, 264–291. [\[CrossRef\]](#)
19. Leukel, J.; González, J.; Riekert, M. Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. *J. Manuf. Syst.* **2021**, *61*, 87–96. [\[CrossRef\]](#)
20. Buddhakulsomsiri, J.; Siradeghyan, Y.; Zakarian, A.; Li, X. Association rule-generation algorithm for mining automotive warranty data. *Int. J. Prod. Res.* **2006**, *44*, 2749–2770. [\[CrossRef\]](#)
21. Jeong, K.; Hong, T.; Chae, M.; Kim, J. Development of a decision support model for determining the target multi-family housing complex for green remodeling using data mining techniques. *Energy Build.* **2019**, *202*, 109401. [\[CrossRef\]](#)
22. Jeong, K.; Hong, T.; Kim, J.; Lee, J. A data-driven approach for establishing a CO2 emission benchmark for a multi-family housing complex using data mining techniques. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110497. [\[CrossRef\]](#)
23. Lv, Z.; Chen, D.; Lv, H. Smart City Construction and Management by Digital Twins and BIM Big Data in COVID-19 Scenario. *ACM Trans. Multimedia Comput. Commun. Appl.* **2022**, *18*, 1–21. [\[CrossRef\]](#)
24. Zhang, H.; Luo, G.; Li, J.; Wang, F.-Y. C2FDA: Coarse-to-Fine Domain Adaptation for Traffic Object Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12633–12647. [\[CrossRef\]](#)
25. Zhang, Y.; Huang, Y.; Zhang, Z.; Postolache, O.; Mi, C. A vision-based container position measuring system for ARMG. *Meas. Control.* **2022**, *56*, 3–4. [\[CrossRef\]](#)
26. Mitici, M.; de Pater, I.; Barros, A.; Zeng, Z. Dynamic predictive maintenance for multiple components using data-driven probabilistic RUL prognostics: The case of turbofan engines. *Reliab. Eng. Syst. Saf.* **2023**, *234*, 109199. [\[CrossRef\]](#)
27. Zhou, G.; Wang, L.; Wang, D.; Reichle, S. Integration of GIS and Data Mining Technology to Enhance the Pavement Management Decision Making. *J. Transp. Eng.* **2010**, *136*, 332–341. [\[CrossRef\]](#)
28. Dindarloo, S.R.; Siami-Irdemoosa, E. Data mining in mining engineering: Results of classification and clustering of shovels failures data. *Int. J. Min. Reclam. Environ.* **2017**, *31*, 105–118. [\[CrossRef\]](#)
29. Moharana, U.; Sarmah, S.; Rathore, P.K. Application of data mining for spare parts information in maintenance schedule: A case study. *J. Manuf. Technol. Manag.* **2019**, *30*, 1055–1072. [\[CrossRef\]](#)
30. Gharoun, H.; Keramati, A.; Nasiri, M.M.; Azadeh, A. An integrated approach for aircraft turbofan engine fault detection based on data mining techniques. *Expert Syst.* **2019**, *36*, e12370. [\[CrossRef\]](#)
31. Gholami, P.; Hafezalkotob, A. Maintenance scheduling using data mining techniques and time series models. *Int. J. Manag. Sci. Eng. Manag.* **2018**, *13*, 100–107. [\[CrossRef\]](#)
32. Kalathas, I.; Papoutsidakis, M. Predictive Maintenance Using Machine Learning and Data Mining: A Pioneer Method Implemented to Greek Railways. *Designs* **2021**, *5*, 5. [\[CrossRef\]](#)
33. Carrasco, J.; López, D.; Aguilera-Martos, I.; García-Gil, D.; Markova, I.; García-Barzana, M.; Arias-Rodil, M.; Luengo, J.; Herrera, F. Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms. *Neurocomputing* **2021**, *462*, 440–452. [\[CrossRef\]](#)
34. Buddhakulsomsiri, J.; Zakarian, A. Sequential pattern mining algorithm for automotive warranty data. *Comput. Ind. Eng.* **2009**, *57*, 137–147. [\[CrossRef\]](#)
35. Hikmawati, E.; Maulidevi, N.U.; Surendro, K. Minimum threshold determination method based on dataset characteristics in association rule mining. *J. Big Data* **2021**, *8*, 1–17. [\[CrossRef\]](#)

**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.