MSIT Management Science and Information Technology
Print ISSN:3079-1006 Online ISSN:3079-1014

Planning Maintenance and Repairs Using the Combination of Data Mining Techniques and Time Series

Mahdi Yousefi Nejad Attari [1]✉, Azin Zahaki [2], Babak Ejlaly [3]

1. Corresponding author, Department of Industrial Engineering, Bon. C., Islamic Azad University, Bonab, Iran. E-mail: 1689597631@iau.ir
2. Faculty of Industrial Engineering, Islamic Azad University Electronic Campus, Tehran, Iran. E-mail: zahakiazin@yahoo.com
3. Department of Industrial Engineering, Faculty of Industrial Engineering and Management, Malek Ashtar University of Technology, Tehran, Iran. E-mail: b.ejlali@mut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received March 19, 2025
Received in revised form July 27, 2025
Accepted August 12, 2025
Available online September 7, 2025

Keywords:
planning
preventive maintenance
time series
clustering
association rules

ABSTRACT

Objective: The primary objective of this study is to develop an effective framework for preventive maintenance and repair planning aimed at reducing maintenance costs, minimizing equipment downtime, and optimizing overtime shifts. By integrating data mining techniques with time series analysis, the research seeks to support managerial decision-making through accurate prediction of repair types, failure times, and operational team performance.

Methods: In this research, clustering algorithms were first employed to classify equipment failures based on similarities in maintenance and repair activities. Subsequently, association rule mining was applied to extract descriptive rules for each cluster and to define effective ranges for influential factors. Time series models were then utilized to predict the time periods during which these factors satisfy the extracted rule ranges. A comprehensive and validated database was constructed using six months of real maintenance records, including failure types, repair operations, equipment downtime, and repair team information.

Results: The results demonstrate that the proposed hybrid approach enables accurate prediction of repair types, equipment downtime, and workforce productivity. By comparing the extracted association rules with time series forecasts, a novel preventive planning mechanism was established. The findings reveal significant differences in the performance of repair teams and indicate that merging or eliminating certain shifts can lead to a substantial reduction in maintenance and repair costs without compromising system reliability.

Conclusion: This study confirms that the integration of data mining techniques and time series analysis provides a robust and practical approach for preventive maintenance and repair planning. The proposed framework enhances resource allocation, reduces unplanned downtime, and supports cost-effective operational strategies. With accurate data collection and continuous database updating, the method can be effectively implemented in manufacturing environments as a decision-support tool for maintenance management.

Cite this article: Yousefi Nejad Attari, M., Zahaki, A., & Ejlaly, B. (2025). Planning Maintenance and Repairs Using the Combination of Data Mining Techniques and Time Series. *Management Science and Information Technology*, *2*(3), 247-268. https://doi.org/ 10.22034/ISS.2025.9172.1035

Publisher: International Scientific Services (ISS).
DOI: https://doi.org/ 10.22034/ISS.2025.9172.1035

1. Introduction

In today's world, data and information are considered as organizational wealth, and large and successful companies and organizations in the world are always looking for more appropriate and commercial use of these virtual resources. On the other hand, with the complexity of business environments, the nature and volume of organizational data has become very different, and an integrated and managerial approach to them becomes necessary. One of the solutions that successful organizations in the world adopt in this field is to create It is a comprehensive data and statistical system and the correct use of data mining and knowledge discovery techniques. The use of big data has been used in various industries to improve supply chain performance (Musah ,2025).

In recent years, many studies have been conducted on preventive maintenance, and a large number of these articles are related to the field of machine learning. The article by Tsallis et al. provides a comprehensive review of the field of applied research on predictive maintenance based on machine learning (Tsallis et al; 2025). Another study presented by Potharaju explores a two-stage machine learning approach for predictive maintenance and anomaly detection in environmental sensor systems (Potharaju et al; 2025). Improving decision-making systems based on artificial intelligence approaches to implement predictive maintenance has been carried out in another study, where this decision-making system was investigated in dynamic conditions (Rajaoarisoa et al; 2025).

Planning maintenance and preventive repairs is also considered one of the most important concerns of manufacturing industries in countries. Considering the development and advancement of industrial machinery manufacturing technology in the world and due to maintenance and repair costs, providing a rational maintenance and repair plan for managers and activists in this field. One of the solutions that can speed up decision-making in such activities and increase the accuracy of decision-making is the correct use of data mining techniques on existing data to discover organizational knowledge or predict the future situation. The infrastructure of an ideal maintenance and repair team includes fixed facilities, including maintenance and repair planning, business and logistics, machinery spare warehouse, mechanical repairs, electrical and electronic repairs, having an experienced and trained repair team. Preventive maintenance and repairs play an important role in the growth of production and productivity of the system, as well as in reducing costs. Therefore, a production company of the automotive industry in East Azerbaijan province is considered, which is responsible for supplying parts such as engine cylinder blocks, crankshafts and bearing caps to the destinations of the parent companies. The examined costs are clear and real, and the time period of the examined data is six months, and the time and type of repairs are known. We are trying to cluster the real data using data mining techniques (Ghaffarian & Bamohabbat, 2022) and extract a series of rules, as well as use collective tree classification algorithms such as Bayes theory and random forest, and compare the results and output rules. Having available the amount of repairs done in the last six months related to this company, first clustering and then extracting association rules and finally using collective tree classification algorithms such as Bayes theory and random forest and the results and output rules We compare Therefore, we can express the main goals of this article based on some data mining techniques (Namjouye Rad & Dadgarpour, 2021) as follows:

- Data clustering in order to sort similar data in a group for statistical analysis
- Extraction of association rules in order to find the relationship for the failure rate of machines
- Tree classification algorithm to obtain the type of repairs for machine breakdowns
- Using the time series algorithm to obtain the efficiency of repair crews.

In the continuation of this article, the theoretical foundations and background of the research are explained in the second part. In the third part, the research method, including the description of the CRISP process and the description of the data, is examined. The fourth part deals with data analysis and application of data mining algorithms as well as time series forecasting. And finally, the fifth part includes the conclusion and description of the findings of the current research.

It should be mentioned that the clustering and extraction of association rules and time series model were used using Clementine, Mini Tab and SPSS software.

2. Theoretical

2.1 Data mining process

Several definitions have been provided for data mining. In some of these definitions, data mining is a tool that enables users to communicate directly with huge amounts of data, and in some more precise definitions, data mining is also considered (Moitra et al., 2021). Based on CRISP data mining standard process model, five phases can be considered for data mining projects, which are: 1) business knowledge 2) data knowledge 3) data preparation 4) modeling 5) evaluation 6) application (Sharma et al., 2017) (Shearer, 2000).

In this standard process, the process does not move linearly, but there is a possibility to go back at every stage, and this process is repeated until we finally get the best result.

Clustering

A cluster is a series of similar data that have similar behavior. It should be mentioned that clustering is the same as classification, with the difference that the classes are not predefined and defined, and the act of grouping data is done without supervision. The act of clustering is such that after the data is placed in the cluster, the relationship of maximum similarity between the members of each cluster and minimum similarity between other clusters is established, that is, the clusters are arranged in such a way that the members within each cluster are the most similar (Ezugwu et al., 2022).

There are different criteria for evaluating clustering algorithms, which can be divided into unsupervised and supervised categories. Unsupervised evaluation criteria, which are sometimes referred to as internal criteria in scientific texts, are those operations that are used according to the information in the data set. The most important task of a clustering algorithm is to minimize the intra-cluster distance and maximize the inter-cluster distance. Therefore, two important factors that are used in all evaluation criteria are: cluster density and cluster separation; And the fulfillment of the objectives of minimizing the intra-cluster distance and maximizing the inter-cluster distance respectively in the group is to maximize the density of each cluster and also to maximize the separation between clusters.

2.2 FP-Growth algorithm

There are many algorithms to generate associative rules, two of the most widely used methods are Apriority and FP-Growth algorithm. In this research, we use the second method. in which it identifies frequent patterns without generating a set of candidate items and with the help of a tree data structure. This algorithm, which follows the divide and conquer strategy, first transforms the data set into a tree called FP-tree. After that, it directly extracts frequent patterns from this tree. In short, it can be said that an FP-tree is a compact representation of transaction set data. To build this tree, the algorithm reads the transaction items one by one and maps them as a path on the tree. Since transactions usually use common items, the paths of transactions overlap on the tree, and for this reason, the tree will be more compact than the original data. In the first scan of the database, the set of items of a member and their support is determined. Also, the set of frequent items are arranged in descending order of their support. Then a tree is built as follows: First, the root of the tree is built with the null tag. After that, the data set is scanned for a second time. The items of each transaction are processed in L order and a branch is created for each transaction. In order to facilitate the navigation of the tree, a table is created in which each item refers to its place in the tree. The tree is complete after scanning all transactions (Shabtay et al., 2021).

2.3 Classification algorithm

Bayes theory: In this method, different categories are considered as a hypothesis with probability. Each new training record increases or decreases the probability of the previous hypotheses being correct, and finally, the hypotheses that have the highest probability are considered as a class and a label is placed on them. This technique deals with classification by combining Bayes theory and causal relationship between data (Schulman, 1984).

A random tree forest produces many decisions. To classify a new object, it places the input vector at the end of each tree in the random forest. Any tree that provides a classification is said to "vote" for that class. The forest selects the classification that has the most votes (among all the trees in the forest) (Hediger et al., 2022).

Research background

Research related to the application of different techniques in maintenance policies can be classified into three groups. The first group of articles were faced with optimization models. Reference (Xia, Jin, Xi, & Ni, 2015) proposed an opportunistic maintenance strategy based on production, considering both machine degradation and batch production characteristics. Reference (Kim et al., 2016) compares condition-based maintenance (CBM) and time-based maintenance (TBM) to compare the optimal combination policy with dynamic programming. Reference (Kumar & Maiti, 2012) has used Fuzzy Analytical Network Process (FANP) to realize effective maintenance policy. Reference (Huang & Chen, 2012) uses data mining methods including K-Mean, Two-Step and C5.0 to classify bridges in several different clusters and draws a decision tree. Reference (Maquee et al., 2012) proposes a method using data mining technique for clustering and extracting rules that predicted the bad situation in a local bus network. In the articles of the second group; The problem of failure mode analysis was addressed. For example, reference (Azadeh et al., 2010) has introduced a timely mechanism and a diagnostic method for machine failure prediction using knowledge acquisition and a fuzzy rule-based inference system. References (Zaluski et al., 2011) and (Gürbüz et al., 2011) using data mining techniques in air bases and aviation bases, identified the important features affecting the amount of warning in the aircraft fleet. In the articles of the third group; Analytical techniques were used to perform equipment prognosis. For example, reference (Gürbüz et al., 2011) used a dynamic evidential reasoning algorithm to accurately predict reliability in turbocharger machines. Reference (Xia, Jin, Xi, Zhang, et al., 2015) presented a real-time rolling gray prediction method to effectively predict machine health. Reference (Ferreiro et al., 2012) presented a maintenance forecasting approach based on a Bayesian network model for both preventive maintenance and maintenance activities. Reference (Xia et al., 2013) proposed a method for multi-level planning based on CBM to predict the need for maintenance. Reference (Deloux et al., 2012) presents a new decision rule to update the environmental information maintenance decision in the behavior of the system state and model the decision framework to use the new information and update the decision.

3.Research Methodology

3.1 Description of the CRISP process

One of the most widely used data mining methods is the CRISP process or the cross standard process (Tounsi et al., 2020) for data mining. As part of its goal to provide more value to its data warehouse customers, America's largest electronic software and hardware company established groups of data mining consultants and technical experts to service its customers' needs. The data mining process in the Crisp data mining standard, according to the knowledge discovery method presented by Osama Fayad and his colleagues (Ho et al., 2005) and (Fayyad et al., 1996), is implemented during six stages. Crisp execution model is a circular and iterative model. Figure 1 shows the flowchart of the research process. In Figure 2, the Crisp process model is drawn cyclically.

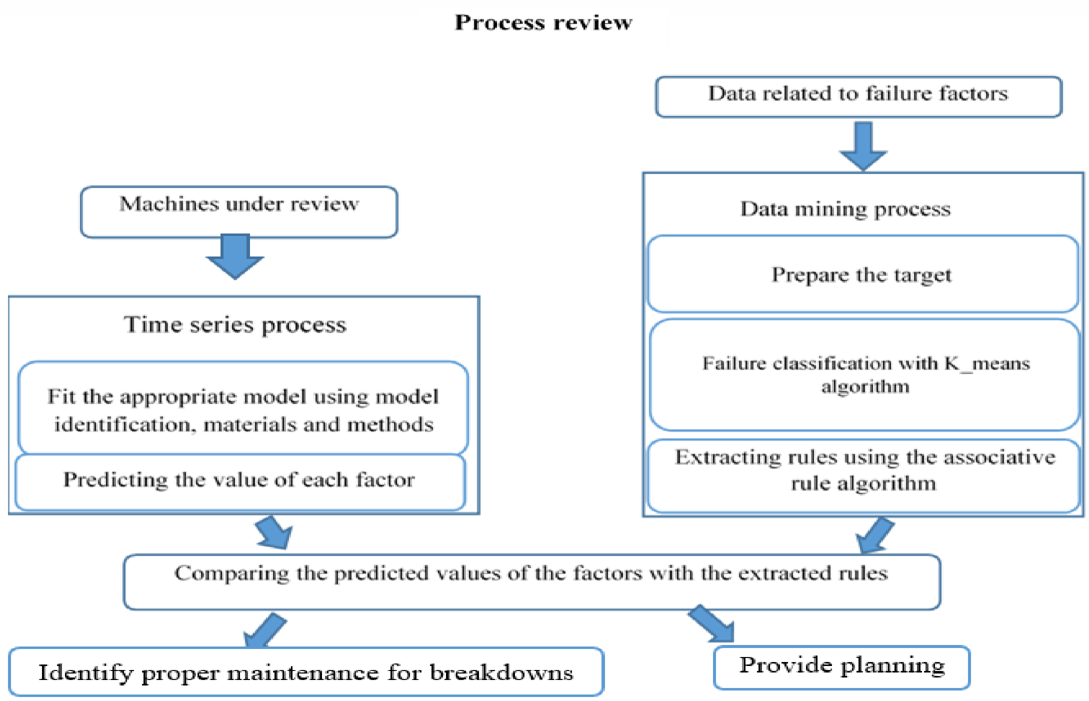


Figure 1. Flowchart of the research objective

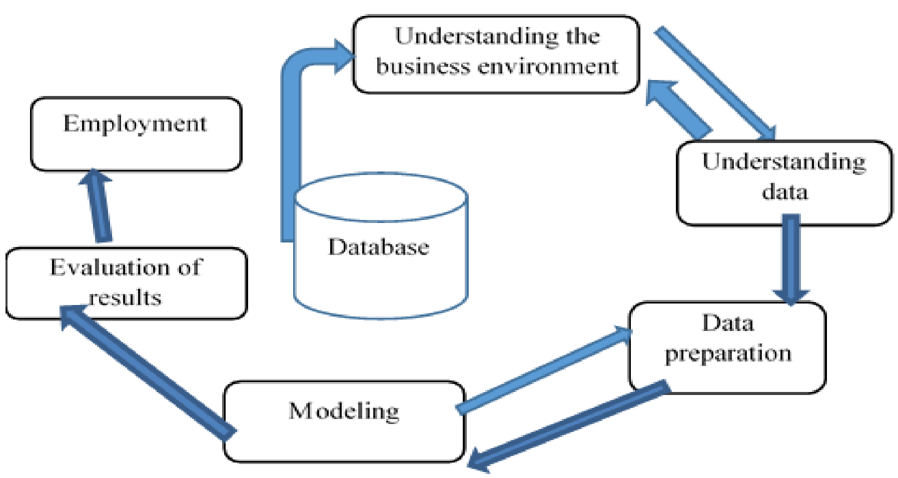


Figure 2. cyclically Crisp process model

3.2 Data description

In this research, we use the real data of an automobile manufacturing company in East Azarbaijan province. The costs under investigation are specific and real. The examined time period of the data related to the period of time is six months. It includes two parts of data collection and data understanding and statistical analysis. The data collection includes 406 pieces of information related to the maintenance and repairs of an automobile parts manufacturing company, which includes cylinder blocks, cylinder heads, crankshafts, and bearing caps. produces. The data contains

6 numerical and discrete columns. The features are defined as described in Table 1, and a part of the raw data is also shown in Table 2 to see how it is collected in Excel software for use in the relevant software.

Table 1. Data table and data type

Title	Variable type
Date of failure	Date
Type of failure	Numerical
Device sleep time	Numerical
Repair team	Nominal
Device type	Nominal
Type of failure	Nominal

Table 2. Sample raw data

Date	type	Time	Team	machine	Parts
1399/06/10	A04	4	B	B01	s2
1399/06/28	T10	2	B	B01	a1
1399/07/20	TO4	20	B	B01	t2
1399/08/11	A06	4	A	B01	s3
1399/08/15	A09	1	C	B01	t3
1399/09/13	T09	6	C	B01	s4
1399/09/25	A14	12	B	B01	s5
1399/10/03	T10	1	A	B01	a1
1399/10/12	A04	2	C	B01	t4
1399/10/21	T06	2	B	B01	s6
1399/11/04	A11	1	B	B01	s7
1399/11/25	T06	1	C	B01	s8
1399/11/29	T07	1	C	B01	c1
1399/12/05	T06	2	A	B01	s9
1399/12/18	T07	1	A	B01	s8
1399/06/19	A11	2	A	B02	s7
1399/06/23	T09	12	B	B02	s4
1399/06/29	T07	2	A	B02	s8
1399/07/19	T06	1	B	B02	s9

3.3 Preprocessing

3.3.1 data cleaning and preparation

In this step, after describing the data with a diagram, if there is missing data, based on the number of missing data, the approach of removing or replacing the missing data is selected. We also check the presence of duplicate lines and additional variables in the software. To validate the data and increase the quality of the raw data of this study, a complete process was carried out to ensure the integrity of the data. This process includes the steps of checking the data compliance with the expected values, evaluating the repeatability of the results, and statistical analysis of deviations. Common techniques include comparison with reference data, statistical tests such as analysis of variance. For example, equipment downtime, which is a numerical characteristic in this study, was determined by the regression method for missing data.

3.3.2 Clustering

To analyze the data, the first step is clustering and grouping the information. To model the analysis of k-clustering, we used means. First, we enter the number of parameters manually and we first determine the optimal number of clusters with the Davis-Bouldin evaluation parameter, and in the analysis stage, we evaluate the most optimal number of clusters with the output of the central table.

3.3.3 Extracting knowledge of association rules

The process of discovering the rules of dependence is one of the important approaches in the modern science of data mining to find rules and patterns in the database, which has received a lot of attention from researchers. In the knowledge extraction step, the goal is to generate association rules from the dataset to extract relationships between features in the field of repairs. The initial hypotheses are as follows: Consider the set I = I1, I2, I3, ... as a set of items. Consider set D as a set of data, i.e. database transactions, such that each transaction contains a set of items, that is, each transaction T is a subset of I (). Each transaction has an identifier called TID. If A is a set of items, we say a transaction T contains A if and only if. An associative law is a proposition in the form of, while, and. We introduce two important parameters for the acceptability of association rules, i.e. support and reliability of the rules. Rule in the set of transactions D has support equal to s if s percent of transactions in D include, and this rule has reliability equal to c if c percent of transactions that include A also include B. Or in other words:

$$\text{Support } (A\Rightarrow B) = P(A\cup B) \quad (1)$$

$$\text{Confidence } (A\Rightarrow B) = P(B|A) \quad (2)$$

The rules that have the lower limit of Support or min-sup and the lower limit of Confidence or min-conf are called strong association rules, and the goal of all algorithms is such rules.

Knowledge extraction (categorization algorithms)

In the classification step, after grouping load data and cluster analysis, we use collective tree classification algorithms such as random forest and simple Bayes probabilistic method and compare the results and output rules. In order to implement the categories, the data should be divided into two categories, training and testing. Using the cross-validation method, the data set is divided into two groups, training and test. In the simple validation method, in all stages of implementation, for example, 70% of the training data set and the rest can be used for testing and calculating indicators; But in a more intelligent way, we use the cross-validation method, which was also used in this research, to have the entire data set once in both training and testing. In this type of validation, the data is divided into K subsets; From these K subsets, each time one is used for validation and another K-1 is used for training. This procedure is repeated K times and all data are used exactly once for training and once for validation. Finally, the average result of these K times of validation is chosen as a final estimate.

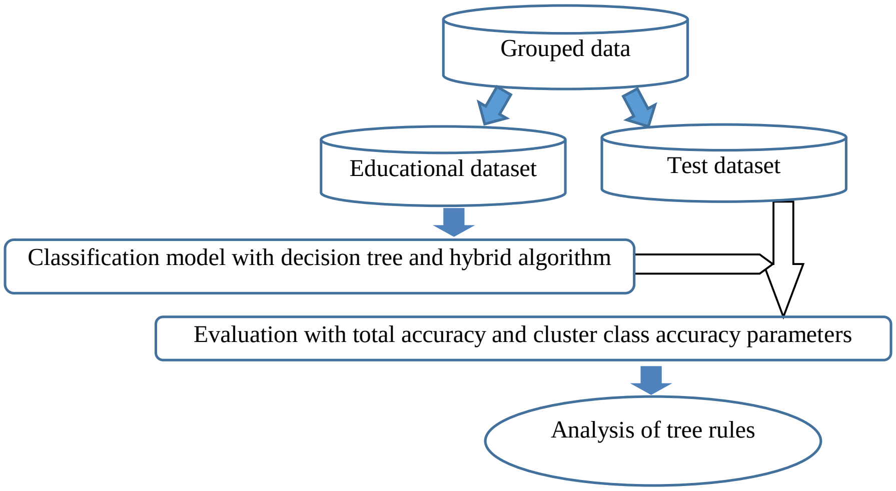


Figure 3. Classification model (knowledge extraction)

3.4 Evaluation criteria of classification algorithms

The confusion matrix shows how the classification algorithm works according to the input data set by different categories of the classification problem. According to Table 3, the concepts of true positive, false positive, true negative and false negative are as follows.

Table 3. Confusion matrix

Estimated Records			
Positive Category (+)	Negative Category (-)		
FP	TN	Negative Category (-)	**Real Records**
TP	FN	Positive Category (+)	

- True positive: This value represents the number of records whose actual category is positive and the category has correctly recognized their category as positive.
- False positives: This value represents the number of records whose actual category is negative and the category classifier incorrectly recognized their category as positive.
- True negative: This value represents the number of records whose real category is negative and the category has correctly recognized their category as negative.
- False negative: This value represents the number of records whose real category is positive and the category classifier incorrectly recognized their category as negative.

The most important criterion is the accuracy or classification rate, and it means that the classifier has correctly classified how many percent of the test records. This accuracy can be calculated based on matrix concepts according to the relation:

$$Accuracy = \frac{Tp + TN}{Tp + TN + Fp + FN} \tag{3}$$

4. Data analysis

4.1 Collecting data

The best method is to observe and record real data, although the number of data is small, but because it is closer to reality, more accurate research and investigation can be obtained and better results can be obtained for a more accurate planning. In this research, a manufacturing company of engine and engine parts and gearbox in the automotive industry has been referred. After the investigations, the amount of six-month failure data, which were almost more accurate, has been recorded and implemented in EXCEL software.

4.2 Data coding

One of the main principles for better execution with data mining software is data coding, for a better planning, you should pay attention to things like: types of failures in the system, number of failures, time of failures of all types of machines. To express this issue better, we must first divide the failures into two parts, mechanical and electrical and electronic. For a better review of these materials, the history of machine breakdowns was extracted and placed in Table 4.

Table 4. The table various failures with the placement of the codes used in the continuation of the work

The title of electronic and electrical failure	Code	The title of mechanical failure	Code
Failure of encoders and rulers	A01	Belt tear	T01
Failure of servo motors	A02	Failure of bearings, roller bearings	T02
Failure of inductive sensors	A03	Failure of couplings and foleys	T03
Failure of relays and contactors	A04	Gear failure	T04
PLC failure	A05	Failure of shafts	T05
Failure of bobbins	A06	Failure of the hydraulic solenoid valve	T06
Failure of power supplies	A07	Oil pressure drop	T07
Failure of micro switches	A08	Rupture of retaining chains	T08
Failure of non-metals and fuses	A09	Failure of hydraulic pumps	T09
Failure of auto transfer motors	A10	Leaking oil connections	T10
Control system failure	A11	Malfunction of the milling table	T11
Display failure	A12	Cut a thorn, pin	T12
Failure of electronic boards	A13	Tool change failure	T13
Malfunction of electric motors	A14	Failure of hydraulic and pneumatic jacks	T14
Loss of machine data	A15	Pallet exchange failure	T15

Table 5 is designed for the devices considered in this research. For any failure, it is possible to solve the problem of device failure by replacing the part or only by repairing and servicing the parts, so for this purpose, a table was also added to describe the repair operations in more detail (Table 6). Finally, the failure table was drawn for each device. SPSS software was used to extract the results of K-means method, decision tree and association rules, and Minitab software was used to extract the results of time series.

Table 1. The name of the machines

Referral code	Device Name	Referral code	Device Name	Referral code	Device Name
M 03	Multi spindle 03	R 02	Radial drill 02	B 01	Boring 01
M 04	Multi spindle 04	F 01	Universal milling cutter 01	B 02	Boring 02
W 01	Werner CNC 01	F 02	Universal milling cutter 02	B 03	Boring 03
W 02	Werner CNC 02	F 03	Universal milling cutter 03	C 01	Daewoo CNC 01
W 03	Werner CNC 03	F 04	Universal milling cutter 04	C 02	Daewoo CNC 02
H 01	Honing 01	F 05	Universal milling cutter 05	C 03	Daewoo CNC 03
H 02	Honing 02	M 01	Multi spindle 01	C 04	Daewoo CNC 04
-	-	M 02	Multi spindle 02	R 01	Radial drill 01

Table 2. Title of repairs

Service description	Error code	Service description	Error code
Lubrication charge	C 2	Wrench connections	a 1
Belt replacement	t 1	Service oriented	s 1
Gear shift	t 2	Relay service	s 2
Replace the fuse	t 3	Clutch adjustment	s 3
Replace the relay	t 4	pump repair	S 4
Bearing replacement	t 5	Electric motor repair	s 5
Hose replacement	t 6	Adjustment of hydraulic valves	s 6
O-ring replacement	t 7	Electric circuit service	s 7
Changing the magazine pin	t 8	Oil pressure adjustment	s 8
Replace the sensor	t 9	Adjust the valve	s 9
Replacing the oven spring	t 10	Fix tool grip	s 10
Fiber disc replacement	t 11	Micro switch setting	s 11
Coupling replacement	t 12	Fibrous disc correction	s 12
Changing the tool pin	t 13	Car guide service	s 13
Replacing the solenoid valve	t 14	Hydraulic jack repair	s 14
Replacing the shaft spur	t 15	Ray Lloyd program	s 15
Packing change	t 16	Ruler service	s 16
Changing table nails	t 17	Sensor adjustment	s 17
Changing the multi-spindle chain	t 18	Modify table chain	s 18
Hydraulic oil change	t 19	Control panel service	s 19
Replacing the micro switch	t 20	Fixed wire failure	s 20
Replacement of connections	t 21	Gearbox failure	s 21
Bearing pin replacement	t 22	Multi-spindle chain service	s 22
-	-	Metalless service	s 23

4.3 K-means clustering

In order to obtain clusters and compare them, different clusters of four, five, six and finally seven clusters have been created. The obtained results are shown in the following figures.

In this cluster, cluster one has the most activity and cluster four has the least activity, and our clustering portfolio is of good quality. Two important factors in this clustering are the ratio of time to the repair team. From this point of view, it is possible to obtain the amount of productivity of the repair team and the sleep time of the device in each breakdown for repairs, and in general, it is possible to say which team with which time. For which repairs. Another point in this clustering is the ratio of the size of the largest cluster to the size of the smallest cluster, which is 15.33 (Figure 4). In this figure, 4 clusters are defined, showing the size of each cluster as a percentage of the total data generated. The influential inputs in this clustering are shown separately for each cluster.

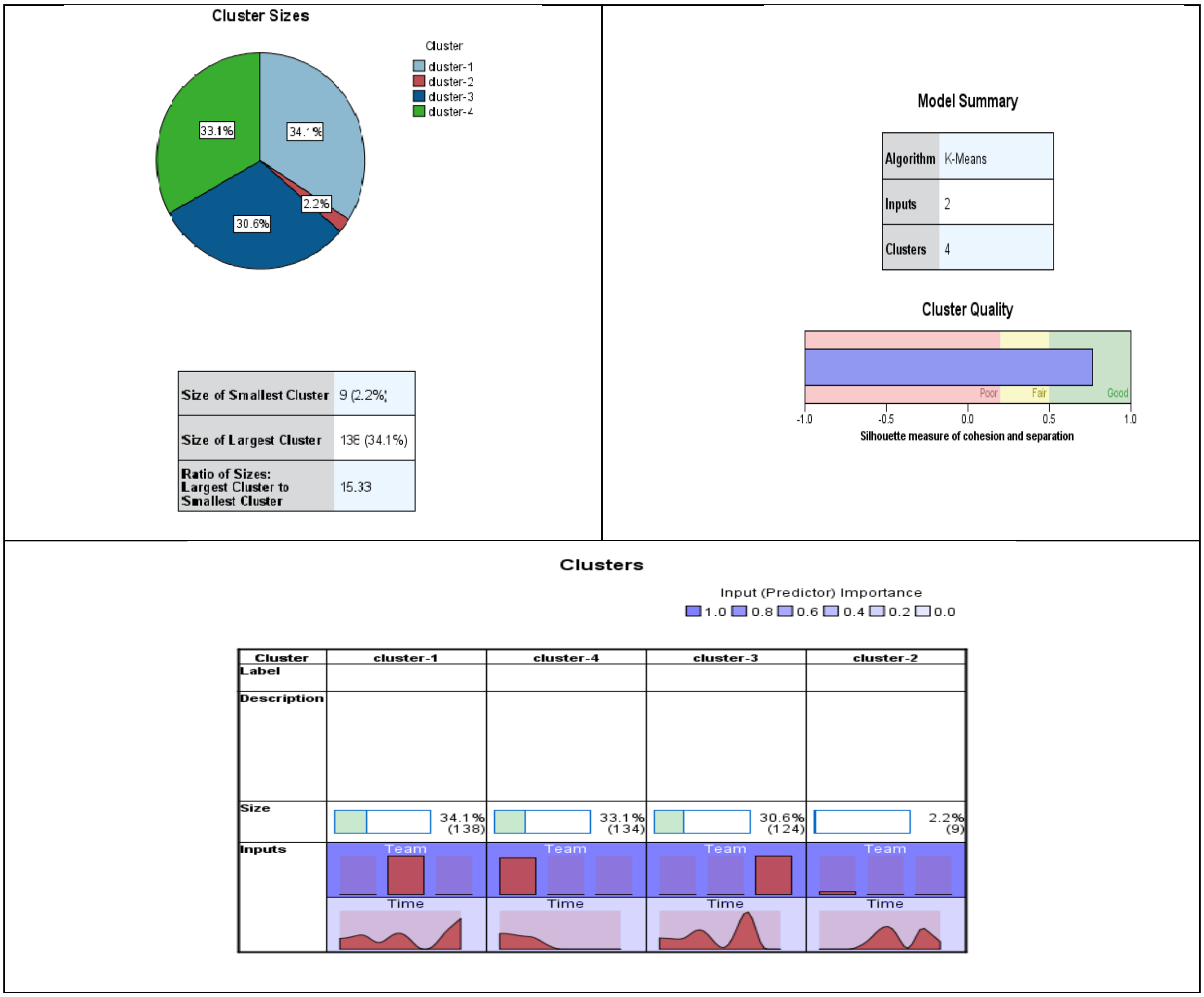


Figure 1. Clustering of four

In the clustering of five, two important factors of time and team have been investigated and this clustering is also of good quality. In this clustering, the amount of team activity has been evaluated in relation to time and failure. The ratio of the size of the largest cluster to the smallest cluster is 45 (Figure 5). The horizontal axis in this figure shows the inputs and the vertical axis shows the frequency of the data.

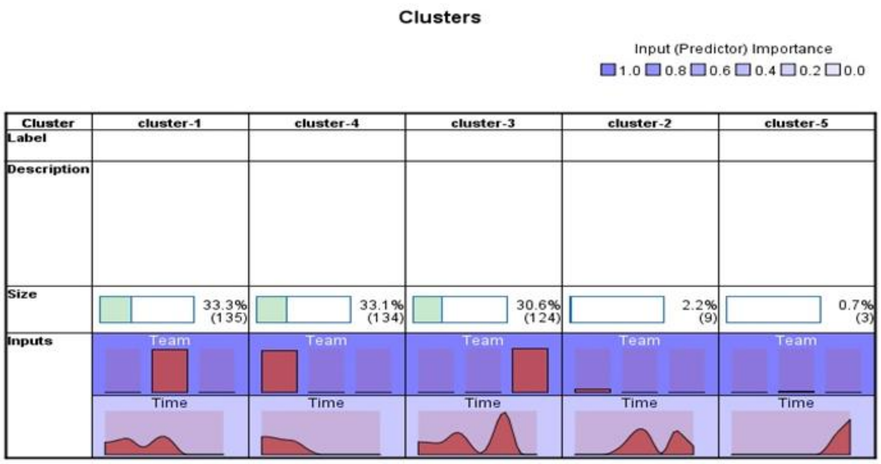


Figure 2. Clustering of five

In the clustering of six, cluster one has achieved the most activity and the quality of clustering is also good, but the ratio of the size of the largest cluster to the smallest cluster has not changed (Figure 6).

The seven-point clustering is also of good quality, and in this clustering, the first cluster has the most activity and the fifth cluster has the least activity, but the ratio of the size of the largest and the smallest cluster has not changed. By comparing the clusters, the result is that there are very few changes in the size and type of clustering, and smaller clusters can be ignored so as not to lose a lot of time for investigation. This indicates correct data and clustering with a good and reliable portfolio. By examining the minor repairs done by each team, it is possible to predict the sleep time of each device during repairs. By referring to the repair list of each team, it can be stated that, for example, to replace a bearing, the device will sleep for 2 hours with team B and one hour with team A, and this information is very important in creating the work schedule of each team. and how much repairs the device needs with this breakdown (Figure 7).

It should be mentioned that figures 4, 5, 6 and 7 show how to cluster. In all these figures the horizontal axis in this figure shows the inputs and the vertical axis shows the frequency of the data.

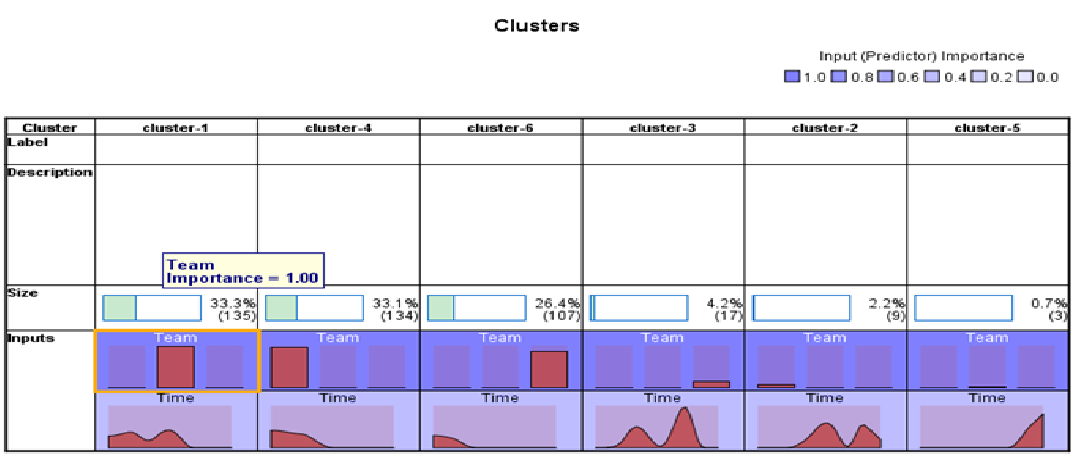


Figure 3. Clustering of six

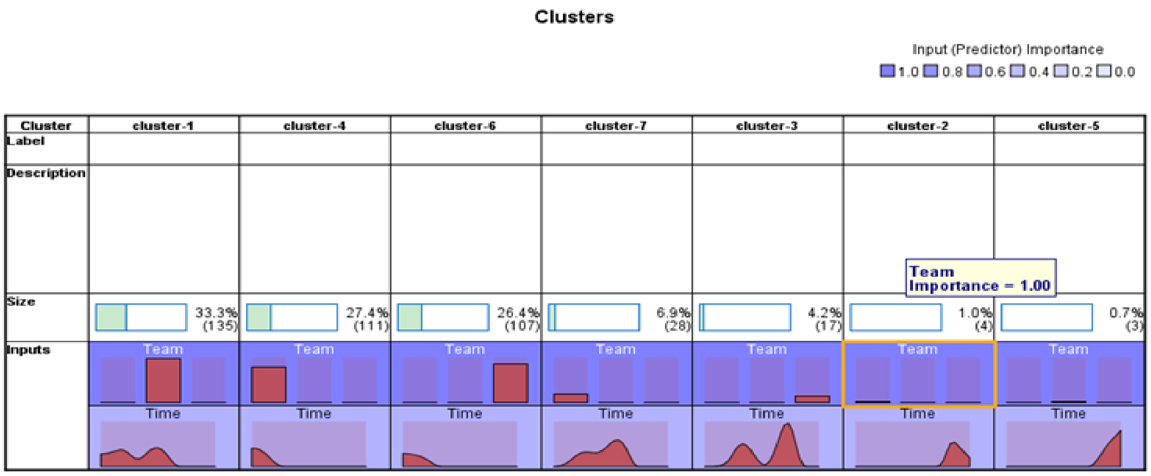


Figure 4. Clustering of seven

4.4 Grouping

For the decision tree, the basis of Part, or repair services that have been performed for various types of failures, is used. One of the most important options for creating a better plan is to predict the type of repairs and the time required for repairs. With such knowledge, it is possible to plan repairs according to the ability of the repair team, the level of sensitivity of the work, and the priority of the production request. In the following, the type of classification and importance of each repair will be explained using the results extracted from the software. In order to obtain a good result in the classification and extraction of suitable nodes, utmost care should be taken in setting the data so that there is no problem in extracting the results. Below you can see the classification of nodes based on the ratio of repairs done to the type of failures.

The first node, the zero node, shows us all the repairs done for all the failures with the frequency type. Most frequency with 8.503% is related to type a1 repairs. So, the first node is formed with its basis. In Figure 8, the number and percentage of all the different states are shown.

The repairs of a1 are related to the failure of T10. By referring to the failure table, we find out that this failure is the leakage of oil connections, and the repair type a1 is wrenching of the connections. In Figure 9, the number and percentage of states at second level are shown.

From node one, we find that T10 failure can be fixed with four types of repairs with a frequency of 12.925% of the total repairs. a1 (wrenching connections) with a frequency of 65.789%, which is the most active, t21 (replacing connections) with a frequency of 5.263%, t6 (replacing hoses) with a frequency of 13.158%, t7 (replacing o-rings) with a frequency of 15.879%, now facing T10 failure is the best decision for a1 repairs.

In the second node, other breakdowns have been investigated. In this node, the most frequent repairs with a frequency of 9.375% of the total repairs related to s9 is the adjustment of the solenoid valve. By referring to the relevant tables, we find that this type of repair is related to T06 failure (solenoid valve failure). So, this next node should be based on T06 failures.

By examining node 5, it is clear that if T06 fails, the first decision for repairs with a frequency of 60% can be S9 (solenoid valve adjustment). Again, in the sixth node, the rest of the failures have been checked based on frequency.

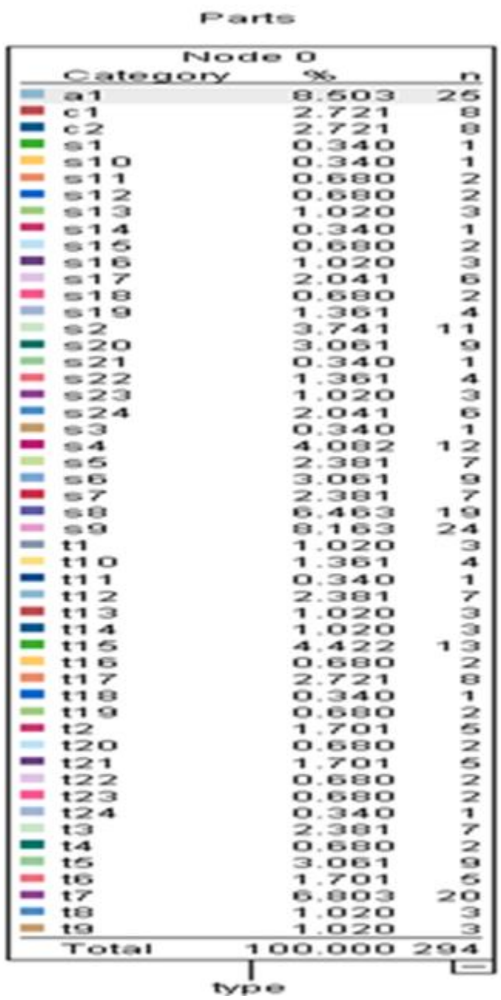


Figure 5. Start node

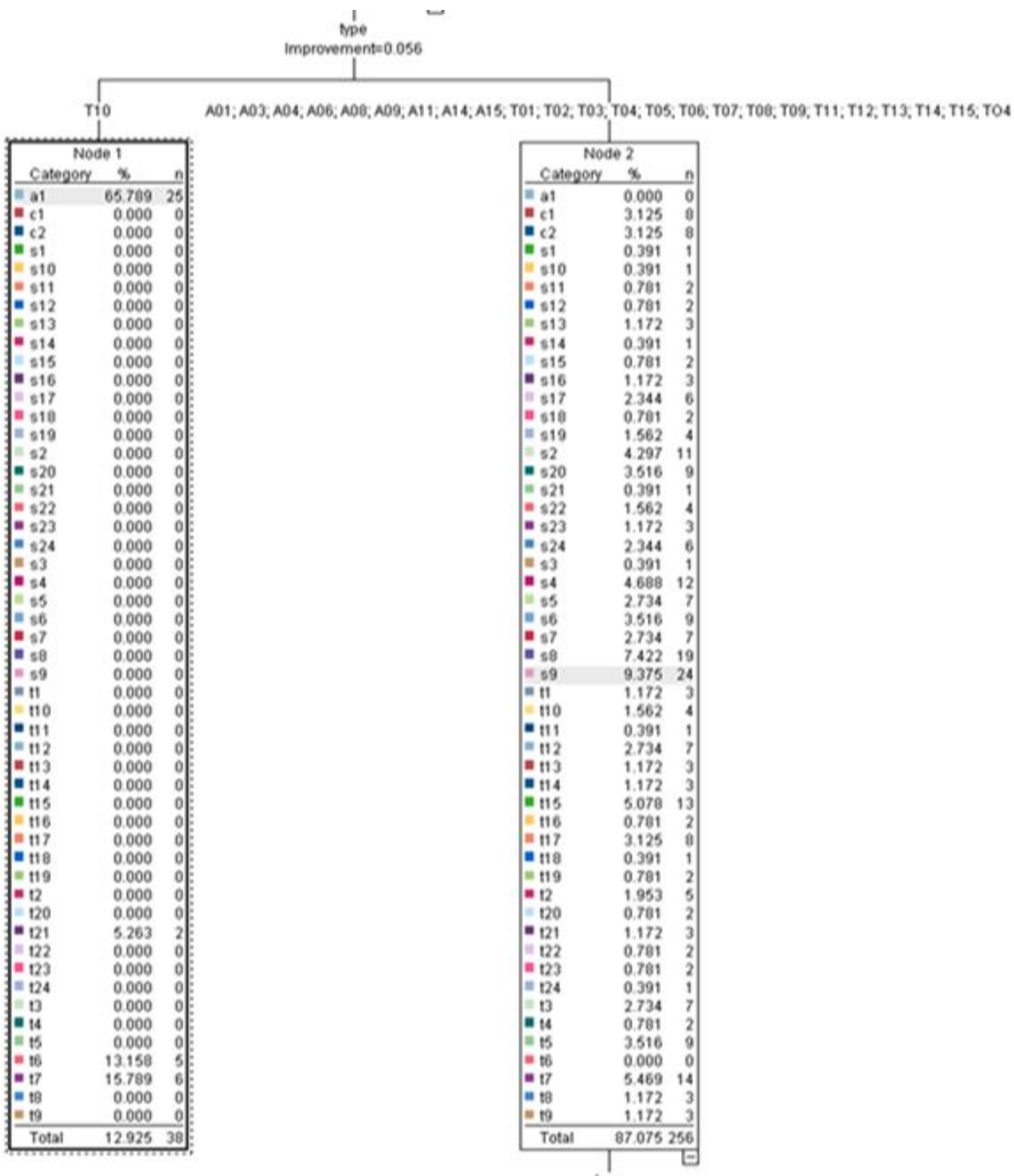


Figure 6. First and second node

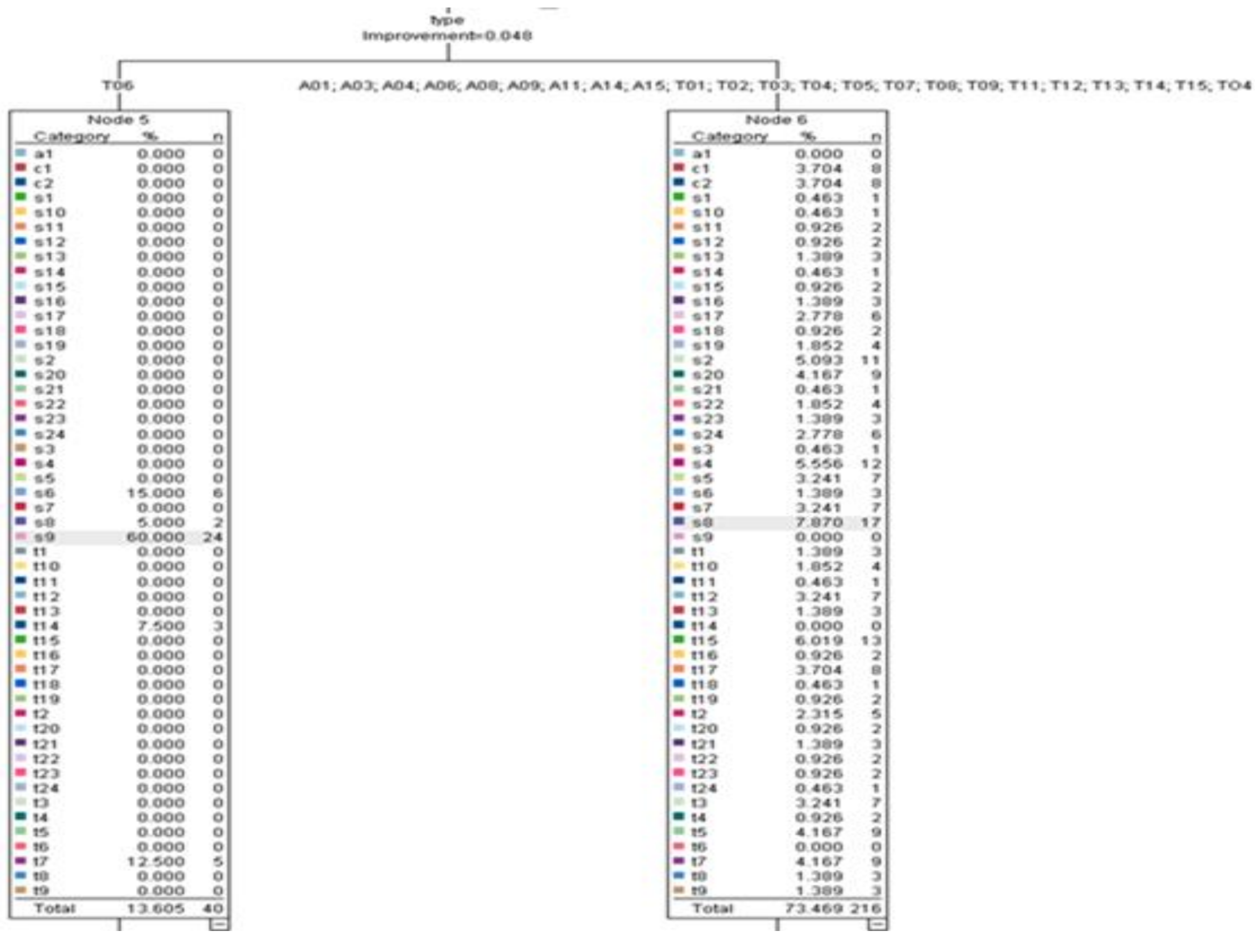


Figure 7. Fifth and sixth nodes

In Figure 10, the number and percentage of states at third level are shown. In this node, it is stated for us that if T06 failure occurs and its sleep time is less than 2.5 hours or equal to 2.5, it is related to s9 repairs, and if it is more than 2.5, it is related to t7 repairs, which of course cannot be done. He relied because the time for different types of repairs is different and it is usually done after the occurrence of a breakdown, in which case the possibility of repairs interfering with other breakdowns cannot be prevented. So, this node is not used for our planning. The general decision tree model is described below.

type in ["T10”] [Mode: a1] => a1

type in ["A01" "A03" "A04" "A06" "A08" "A09" "A11" "A14" "A15" "T01" "T02" "T03" "T04" "T05" "T06" "T07" "T08" "T09" "T11" "T12" "T13" "T14" "T15" "TO4”] [Mode: s9]

type in ["T06”] [Mode: s9]

Time <= 2.500 [Mode: s9] => s9

Time > 2.500 [Mode: t7] => t7

type in ["A01" "A03" "A04" "A06" "A08" "A09" "A11" "A14" "A15" "T01" "T02" "T03" "T04" "T05" "T07" "T08" "T09" "T11" "T12" "T13" "T14" "T15" "TO4”] [Mode: s8]

type in ["T09”] [Mode: s4] => s4

type in ["A01" "A03" "A04" "A06" "A08" "A09" "A11" "A14" "A15" "T01" "T02" "T03" "T04" "T05" "T07" "T08" "T11" "T12" "T13" "T14" "T15" "TO4”] [Mode: s8]

type in ["T07”] [Mode: s8] => s8

type in ["A01" "A03" "A04" "A06" "A08" "A09" "A11" "A14" "A15" "T01" "T02" "T03" "T04" "T05" "T08" "T11" "T12" "T13" "T14" "T15" "TO4”] [Mode: t15]

type in ["A04”] [Mode: s2] => s2

type in ["A01" "A03" "A06" "A08" "A09" "A11" "A14" "A15" "T01" "T02" "T03" "T04" "T05" "T08" "T11" "T12" "T13" "T14" "T15" "TO4”] [Mode: t15] => t15

4.5 Association rules

In this model, considering the two factors of failure and repairs, three rules have been created for us. Table 3. Established community rules

Table 7. Established community rules

Consequent	Antecedent	Confidence %
Parts = a1	type = T10	68.0
Parts = s9	type = T06	60.78431372549019
Parts = s8	type = T07	44.0

By examining the above rules (Table 7), we found that if a T10 failure occurs, a1 repairs will be performed with a 68% probability, and in case of a T10 failure, S9 repairs will be performed with a 68% probability, and with a T07 failure, s8 repairs will be performed with a 44% probability. % will be done.

4.6 Time series

One of the most important issues that can help with preventive maintenance is the estimation of machine downtime. This time is very effective for planning maintenance teams. With sensitivity analysis on the parameters of the time series analysis method, the length of 72 was more accurate for the moving average. In this model, the duration of the repairs performed by the repair crews during the sleep time of the device is used. That is, what is the amount of repair time to repair a breakdown and it will predict for us how much time each team needs for repairs in the next hours.

TEAM A Moving Average for Time

Data Time

Length 143

N Missing 0

Moving Average

Length 72

Accuracy Measures

MAPE 58.5787

MAD 2.4175

MSD 19.8853

Forecasts

Period	Forecast	Lower	Upper
144	4.34722	-4.39283	13.0873
145	4.34722	-4.39283	13.0873
146	4.34722	-4.39283	13.0873
147	4.34722	-4.39283	13.0873
148	4.34722	-4.39283	13.0873

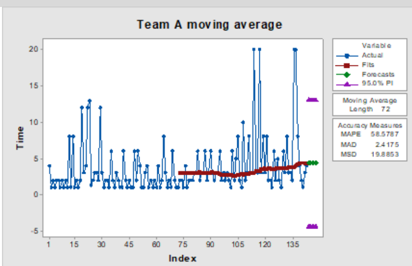
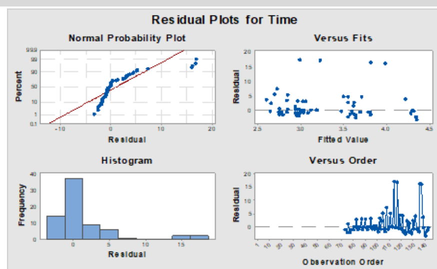
Moving Average Plot for Time	Residual Plots for Time
	

Figure 8. The results of TEAM A

This forecast is considered for the next seventy-two hours and 5 forecasts have been requested. For team A, 4.34 hours have been predicted for breakdowns in the coming hours, which means that on average this team needs 4.3 hours for each breakdown. As it is clear in the first diagram, the level of operation is acceptable and the next diagram shows the amount of errors for this prediction. Considering that our data is around the horizontal axis and close to the base axis, it shows real data and this error rate is normal. In the histogram chart, it can be seen that the most activity is done at the zero point, and the further away from this point, the amount of errors decreases. The results and graphs of the next teams can also be seen below. Figures 11, 12 and 13 also show these results. In all three figures, the horizontal axis represents index and the vertical axis represents time of future events for each team.

Data TIME

Length 138

NMissing 0

Moving Average

Length 72

Accuracy Measures

MAPE 123.988

MAD 1.989

MSD 5.713

Forecasts

Period	Forecast	Lower	Upper
139	2.66667	-2.01790	7.35123
140	2.66667	-2.01790	7.35123
141	2.66667	-2.01790	7.35123
142	2.66667	-2.01790	7.35123
143	2.66667	-2.01790	7.35123

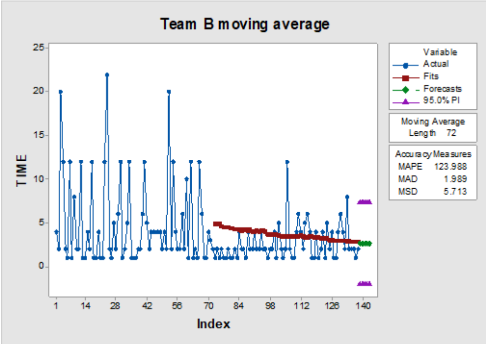


Figure 9. TEAM B survey results

TEAM C Moving Average for Time **Moving Average Plot for Time**

Data TIME

Length 124

NMissing 0

Moving Average

Length 72

Accuracy Measures

MAPE 110.431

MAD 2.011

MSD 6.041

Forecasts

Period	Forecast	Lower	Upper
125	2.76389	-2.05354	7.58132
126	2.76389	-2.05354	7.58132
127	2.76389	-2.05354	7.58132
128	2.76389	-2.05354	7.58132
129	2.76389	-2.05354	7.58132

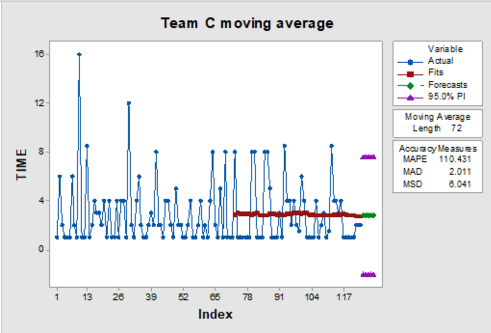


Figure 10. The results of TEAM C

5. Conclusion and Management suggestions

This research has attempted to address the challenges of implementing preventive maintenance in industrial environments. One of the most important challenges is the lack of use of generated data in current maintenance systems. This data can provide a suitable framework for making decisions and planning proper maintenance. Data analysis in this area can provide the necessary basis for predicting machine downtime and repair times.

After examining the results, it was concluded that by using data mining with the integration of time series (Abbaspour Ghadim Bonab, 2022), it is possible to make a proper planning to prevent breakdowns and carry out the necessary repairs on time, preventing machines from sleeping and stopping production. and incur additional repair costs and the requirement to have additional shifts to compensate for the time required for machinery repairs. As seen in the results, with this method, in the event of a failure, it is possible to easily predict the amount of sleep time of the device and the type of repairs needed to eliminate the failure, and by examining the time series, in addition to the average amount of time for the repairs of each The team obtained the productivity of the teams. If you look closely at the results of the time series, you will see that team B had the least performance and team A had the most productivity. Therefore, by removing the shift of maintenance team B and integrating the employees of this team as needed in other teams, the cost of one work shift can be reduced and the productivity of other teams can be increased. In this method, with the lowest cost and with an accurate database and accurate recording of all failures and operations, the exact time of the failure and the completion of repair operations and the items and tools needed to replace the defective parts or the required time are known. To make spare parts, to create a complete spare warehouse, and to obtain the time of upcoming failures by using an up-to-date database for an accurate planning of failure prediction. As seen in the results obtained in the decision tree, in this method it is easy to predict the type of repairs at the time of failure and take action as soon as possible to carry out the repairs or provide the required parts and save the time and costs.

By having detailed repair monitoring information, according to the results of association rules, it is confirmed that by observing which failure, what kind of repairs can be considered for the device. So, the final result obtained from this article leads us to do a detailed planning, but it should be kept in mind that the entire function of this method is achieved by having a complete and updated database, so accurate registration The information and events that took place in the field of preventive maintenance and repairs are very important and sufficient care should be taken in this matter. The best way to record this information is to use Excel software, which with the facilities available in this program, can be created by creating tables for each of the machines separately, like the sample tables in this article for each of the important machines. Or the so-called bottleneck in the production line provided separate planning and achieved a more precise process than any of the mentioned machines.

The management suggestions resulting from the findings of this research can be expressed as follows:

- Although this research was conducted on a manufacturing company, the research process of this article in order to reduce repair time and costs by planning preventive maintenance and repairs can be carried out in other manufacturing centers with the same process. In other companies, the type of repairs required will vary, but the way data is collected and analyzed can achieve the necessary results.
- Cost reduction is stated as a main objective in this research. However, the costs are based on the reduction of machine downtime using data analysis. Estimating this time reduction requires using the results of this research. Therefore, until this research is implemented in a real environment, it is not possible to talk about the quantitative cost reduction.

This research uses a clustering algorithm to cluster repair types and an association rules algorithm to define rules related to repairs, and no other data analysis techniques have been considered. Researchers can also use combined techniques in this field.

The use of other machine learning techniques that predict preventive maintenance planning along with estimated repair times and machine downtime is recommended in future research. These techniques could include:

- Bayesian networks
- Random forest algorithms

References

Abbaspour Ghadim Bonab, A. (2022). A comparative study of demand forecasting based on machine learning methods with time series approach. Journal of applied research on industrial engineering, 9(3), 331-353. https://doi.org/10.22105/jarie.2021.246283.1192

Azadeh, A., Ebrahimipour, V., & Bavar, P. (2010). A fuzzy inference system for pump failure diagnosis to improve maintenance process: The case of a petrochemical industry. Expert Systems with Applications, 37(1), 627-639. https://doi.org/10.1016/j.eswa.2009.06.018

Davies, D. L., & Bouldin, D. W. (2009). A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, (2), 224-227. https://doi.org/10.1109/TPAMI.1979.4766909

Deloux, E., Castanier, B., & Bérenguer, C. (2012). Environmental information adaptive condition-based maintenance policies. Structure and Infrastructure Engineering, 8(4), 373-382. https://doi.org/10.1080/15732479.2011.563095

Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. Engineering Applications of Artificial Intelligence, 110, 104743. https://doi.org/10.1016/j.engappai.2022.104743

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34. https://doi.org/10.1145/240455.240464

Ferreiro, S., Arnaiz, A., Sierra, B., & Irigoien, I. (2012). Application of Bayesian networks in prognostics for a new Integrated Vehicle Health Management concept. Expert Systems with Applications, 39(7), 6402-6418. https://doi.org/10.1016/j.eswa.2011.12.027

Ghaffarian, H., & Bamohabbat, A. (2022). Classification and Prediction of Customer Categories Using Combination of LRFM Method, Quartiles and Multi-Class Data Mining Methods. Karafan Quarterly Scientific Journal. https://doi.org/10.48301/kssa.2022.316104.1852

Gürbüz, F., Özbakir, L., & Yapici, H. (2011). Data mining and pre-processing application on component reports of an airline company in Turkey. Expert Systems with Applications, 38(6), 6618-6626. https://doi.org/10.1016/j.eswa.2010.11.076

Hediger, S., Michel, L., & Näf, J. (2019). On the use of random forest for two-sample testing. arXiv preprint arXiv:1903.06287. https://doi.org/10.48550/arXiv.1903.06287

Ho, T. B., Cheung, D., & Liu, H. (2005). Advances in knowledge discovery and data mining. 9th Pacific-Asia Conference, Vietnam. https://doi.org/10.1007/3-540-45357-1

Huang, R. Y., & Chen, P. F. (2012). Analysis of influential factors and association rules for bridge deterioration using national bridge inventory data. Life-Cycle and Sustainability of Civil Infrastructure Systems: Proceedings of the Third International Symposium on Life-Cycle Civil Engineering (IALCCE'12), Vienna, Austria, October 3-6, 2012, 115. https://doi.org/10.51400/2709-6998.1812

Kim, J., Ahn, Y., & Yeo, H. (2016). A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy. Structure and Infrastructure Engineering, 12(12), 1525-1536. https://doi.org/10.1080/15732479.2016.1149871

Kumar, G., & Maiti, J. (2012). Modeling risk based maintenance using fuzzy analytic network process. Expert Systems with Applications, 39(11), 9946-9954. https://doi.org/10.1016/j.eswa.2012.01.004

Maquee, A., Shojaie, A. A., & Mosaddar, D. (2012). Clustering and association rules in analysing the efficiency of maintenance system of an urban bus network. International Journal of System Assurance Engineering and Management, 3(3), 175-183. https://doi.org/10.1007/s13198-012-0121-x

Musah, L. (2025). Enhancing Supply Chain Performance and Agility in the Healthcare Industry Through Big Data Analytics: A Complex Adaptive Systems Theory Approach. Management Science and Information Technology, 2(1), 1-19. https://doi.org/10.22034/ISS.2025.8740.1025

Moitra, A. K., Bhattacharya, J., Kayal, J. R., Mukerji, B., & Das, A. K. (Eds.). (2021). Innovative exploration methods for minerals, oil, gas, and groundwater for sustainable development. Elsevier. https://doi.org/10.1016/C2020-0-00590-6

Namjooye, R. A. A., & Dadgarpour, M. (2021). Detection of network penetration by data mining and using machine learning via SVM algorithm. https://sid.ir/paper/379238/en

Potharaju, S., Tirandasu, R. K., Tambe, S. N., Jadhav, D. B., Kumar, D. A., & Amiripalli, S. S. (2025). A two-step machine learning approach for predictive maintenance and anomaly detection in environmental sensor systems. MethodsX, 14, 103181. https://doi.org/10.1016/j.mex.2025.103181

Schulman, P. (1984). Bayes' theorem—a review. Cardiology clinics, 2(3), 319-328. https://doi.org/10.1016/S0733-8651(18)30726-4

Shabtay, L., Fournier-Viger, P., Yaari, R., & Dattner, I. (2021). A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data. Information Sciences, 553, 353-375. https://doi.org/10.1016/j.ins.2020.10.020

Sharma, V., Stranieri, A., Ugon, J., Vamplew, P., & Martin, L. (2017, May). An agile group aware process beyond CRISP-DM: a hospital data mining case study. In Proceedings of the International Conference on Compute and Data Analysis ,109-113. https://doi.org/10.1145/3093241.309327

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. Journal of data warehousing, 5(4), 13-22. https://doi.org/10.4236/jss.2015.310021

Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R. A. (2025). Application-Wise Review of Machine Learning-Based Predictive Maintenance: Trends, Challenges, and Future Directions. Applied Sciences, 15(9), 4898. https://doi.org/10.3390/app15094898

Tounsi, Y., Anoun, H., & Hassouni, L. (2020, March). CSMAS: Improving multi-agent credit scoring system by integrating big data and the new generation of gradient boosting algorithms. In Proceedings of the 3rd international conference on networking, information systems & security,1-7. https://doi.org/10.1145/3386723.338785

Xia, T., Jin, X., Xi, L., & Ni, J. (2015). Production-driven opportunistic maintenance for batch production based on MAM–APB scheduling. European Journal of Operational Research, 240(3), 781-790. https://doi.org/10.1016/j.ejor.2014.08.004

Xia, T., Jin, X., Xi, L., Zhang, Y., & Ni, J. (2015). Operating load based real-time rolling grey forecasting for machine health prognosis in dynamic maintenance schedule. Journal of Intelligent Manufacturing, 26(2), 269-280. https://doi.org/10.1007/s10845-013-0780-8

Xia, T., Xi, L., Zhou, X., & Lee, J. (2013). Condition-based maintenance for intelligent monitored series system with independent machine failure modes. International Journal of Production Research, 51(15), 4585-4596. https://doi.org/10.1080/00207543.2013.775524

Zaluski, M., Létourneau, S., Bird, J., & Yang, C. (2011). Developing data mining-based prognostic models for cf-18 aircraft. https://doi.org/10.1115/1.4002812