Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Developing a Model for Product Quality Prediction and Improvement Using the Decision Tree Algorithm and Data Envelopment Analysis – a Case Study: Manufacturers of the Tiba Anti-roll Bar in Iran

Nadereh Sadat Rastghalam[1], Roya Mohammad Alipour Ahari[2*], Ahmad Reza Shekarchizadeh[3], Atefeh Amindost[4]

[1]Department of Industrial Engineering,Najafabad Branch, Islamic Azad University,Najafabad,Iran(nadereh_86@yahoo.com)

[*2]Department of Industrial Engineering,Najafabad Branch, Islamic Azad University,Najafabad Iran.(roya.ahari@gmail.com)

[3]Department of Management,Najafabad Branch, Islamic Azad University,Najafabad,Iran (ahmad_shekar2@hotmail.com)

[4]Department of Industrial Engineering,Najafabad Branch, Islamic Azad University,Najafabad,Iran.(atefeh_amindoust@yahoo.com)

Abstract

Parallel to technological advancements, industries are paying closer attention to cost-effectiveness and product quality as prerequisites of good performance in a competitive world. Using the tools of artificial intelligence and machine learning can cut down on costs and wastes while increasing product quality. The present study employed the decision tree algorithm to explore waste patterns in the production line of anti-roll bars in Tiba. To do this, first, a database comprised 4169 pieces in the form of nine characteristics, and then one class was formed. After determining the patterns, the rules were evaluated via data envelopment analysis. By exploring the most important rules and finding solutions to eliminate them, the quality of the final product was foreseeably enhanced, consequently reducing the percentage of waste pieces and reworking. The proposed approach is recommended for companies with a high rate of waste and different working stations. According to the results, the most important criteria affecting the breakdown include the quality of cooling and soldering. In addition, the accuracy of the C5 algorithm was 94% in predicting the piece quality. The present study evaluated four rules at a depth of 1, 12 rules at a depth of 2, and 8 rules at a depth of 3. By exploring and evaluating such rules, the waste and reworking reduced while the quality increased. The model was validated by implementing reforming priorities in rolling and soldering from April to September 2021, suggesting that when the conditions were sustainable and the quality of input materials as well as other variables remained unchanged in 2021, product quality improved by 7%. This shows that the model is appropriately valid.

Keywords: Assessment, Data envelopment analysis, Decision tree, Quality, Wastes.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

1. Introduction

Considering the important role of quality and reducing costs in production, companies and organizations are more inclined to use quality control tools (Podrzˇaj, Simoncˇicˇ; 2011). Using statistical quality control tools would increase productivity, prevent the production of defective pieces, counter unfavorable settings and occurrences, and increase awareness in all parts of the manufacturing industry (Andersen et al; 2011) According to frequent experiments, production units that benefit from statistical quality control methods enjoy considerable efficiency (Bouslah et al;2016) On the other hand, although factories and industries in Iran welcome statistical quality control methods to improve product quality, more reliable instruments are needed to control the statistical quality control processes, especially since the volume of data is high. For this reason, data mining tools are increasingly used for improving the process of statistical quality control. For instance, Devi et al. (2018) examined effective factors on water quality using a pie chart, Scott plot, and clustering. Loukas (2018) used Bayesian Algorithm to examine phone call quality and proposed relevant solutions. Bordin et al. (2017) presented a model by linear programming to analyze battery failure in solar power systems. Considering the extensiveness of the data mining algorithms and their capabilities in exploring the rules, it is worthwhile to visualize how data mining instruments can improve and increase the process of quality control and how a method can recognize defective pieces before the product is completed, thereby preventing the production of defective pieces. As statistical techniques cannot help in this regard, the present study proposed an approach to discover defective pieces before they are produced. To do so, first, a defective database was formed and, after collecting the data, the decision tree algorithm was used for determining the accuracy of predicting product quality in the pieces. Then, by data envelopment analysis, the extraction rule was assessed and the working stations were evaluated by relevant rules that verified each station. One of the reasons that this model is used for improving product quality and discovering the rules is that the variables in the production line of the working station are non-qualitative, meaning that the variables have a nonlinear effect. Thus, it is recommended that manufacturers use other algorithms that have no defined prerequisite and lack prediction capabilities based on special conditions. Moreover, using techniques such as DOE (Design of Experiments) cannot contribute much because the multiplicity of the factors involved can make the experiments faulty. For instance, if there are 8 working stations, there must be 8 different tests so that different modes are examined and, if each station has more than two modes, the number of modes increases even more, thereby making this technique impracticable.

Accordingly, in the present study, independent variables comprise data that pertain to the statistical quality control process at different stages of production, from the first working station to the last working station. Dependent variables include the quality of the pieces at the final test stage of quality control. In taking advantage of an optimal decision tree algorithm, the present study aimed to determine the patterns that lead to the breakdown of pieces. These patterns were assessed and the characteristics affecting them were determined.

Regarding the outline of this article, first, the theoretical concepts of the research are described as data mining, decision tree algorithm, and data envelopment analysis. Then, the research

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

background is reviewed, the research method is presented, and the experiment is designed. Ultimately, the results are analyzed and discussed to make relevant conclusions and present recommendations.

2. Literature review

2.1. The background

Industries frequently use statistical quality control instruments to improve the quality of their products. As the volume of data is high, more reliable instruments are required to control the processes of statistical quality control. Thus, many studies have employed data mining to reinforce the statistical quality control processes. For instance, Hang et al. (2020) examined data mining in the construction industry while considering the prevailing situation, opportunities, and future prospects. They pointed out that the construction industry has a considerable growth in data production. In the construction industry, data mining (DM) usually emerges from a significant amount of data and serves as an important tool for the discovery of knowledge. Despite the considerable growth of DM programs in the construction industry, systematic explorations of DM programs are less existent. Thus, the present study makes a comprehensive review of the literature on DM programs, characterized by the construction industry, that has been published from 2001 to 2019. DM applications in the construction industry are becoming increasingly popular, especially after 2016, owing much of their growth to Chinese manufacturers. In this article, the main sources of data, DM functions, and frequently-used DM techniques in the construction industry have been discussed in detail. Nine main fields of application have been identified and their main interests were found to be focused on various energy dimensions, safety management, building occupancy, occupant behavior, material performance, and knowledge discovery.

Using relevant research findings, four main challenges and four directions were presented for future research. The present study provides a comprehensive understanding of the most advanced DM applications and heuristic implications for future research .
Devi et al. (2018) carried out research on predicting water quality, using R software to generate analyses for predicting the quality of mineral water. In the said research, tools such as pie charts, Scott plots, and clustering were used for investigating the factors that make high or low qualities. The research concluded that by using the mentioned instruments, the product quality can be predicted properly and to a large extent . Loukas (2018) predicted the quality of phone calls and offered solutions to improve them using Bayesian network algorithm. While using various variables, the researchers predicted customer satisfaction of the phone calls, with the rules leading to high or low satisfaction. Bordin et al. (2017) presented a model using linear programming models for analyzing battery deficiencies in solar electric systems. The goal was to replace new batteries to prevent increasing repair costs. Accordingly, a linear programming model was formed. Baysal et al. (2017) conducted a statistical review using the decision-making tree analysis on previous publications that described corrections of catalyst vapor to develop research methods and

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

enhance performance. Accordingly, a database was formed that included 5508 experimental data sites for correcting methane vapor using 81 research articles, out of an initial number of 453 articles that were reviewed, published between 2004 and 2014. These data were analyzed by decision tree algorithms and the rules were extracted. Alves et al. (2017) investigated linear programming and genetic algorithm for the production of hydrothermal plants and their maintenance planning. It presented a hybrid method for solving production maintenance planning. The proposed mathematical model was run and a better solution was to reduce costs by 10.72% less than that of the base plan. Kamikawaji et al. (2016) conducted research to design a decision tree based on functional attributes in stochastic conditions for error detection in data observed from the ocean because, for climatic predictions, oceanic data are used. Tseng et al. (2016) conducted research on electronic quality control via the support vector machine (SVM) method. Sensor-based, automatic, and computerized inspection methods replaced traditional quality control based on sampling techniques. The production equipment were connected to the network and the status of machinery was monitored. Visual sensors measured the dimensions and recorded all data in the system. The analysis generated results via computer algorithms . Huang et al. (2016) dealt with the prediction of sales using data mining techniques. They found that the accuracy of the presented model was higher than that of a single model. Sellers can use the proposed system for precise predictions of sales from different products. Table 1 summarizes the research.

Table 1. The summary of the Conducted Studies

Row	Study	The summary of the model
1	Hang yan et al. (2020)	Data mining in the construction industry
2	Devi et al. (2018)	Predicting water quality with instruments such as pie charts, Scott plots, and clustering
3	Loukas (2018)	Predicting phone call quality and solutions to improve them by Bayesian Algorithm
4	Bordin et al. (2017)	Using mathematical linear programming models to analyze battery failure in the solar power
5	Baysal et al. (2017)	Analyzing the use of a decision tree to modify methane vapor
6	Kamikawaji et al. (2016)	Predicting weather with ocean data by a decision tree
7	Tseng et al. (2016)	Using sensor-based investigation methods
8	Huang et al. (2016)	Using data mining to predict sales

2.2. Data mining

Data mining is the process of searching and discovering different models, summarizing, and obtaining values from specific data sets. Data mining includes selecting and using computer-based

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

tools to solve current issues and obtain automated solutions. Data mining is not a random application of statistical methods, machine learning, and instruments. Moreover, data mining does not randomly apply different analytical methods, but is a new phenomenon, a decision-making process with programmed accuracy which is highly useful, promising, and appropriate. One of the efficient instruments in data mining is the decision tree algorithm. The decision tree is an efficient and unique method for ranking and classifying data. The decision tree extensively uses a logical method. There are a good number of inductive algorithms in the decision tree that are mainly used in machine learning and are described in the literature on applied statistics. They provide learning methods that create decision trees from sets of input-output samples. A sample training system of the decision tree applies a top-to-bottom approach which creates a solution in one part of the search space. This method guarantees that a simple, but not necessarily the simplest, tree will be found. A decision tree includes nodes with tested features. The external branches of a node are in line with all possible outputs of that test in the node (Kim et al; 2014) .

The most important part of the C4.5 algorithm is when an initial decision tree is created from a set of training samples. As an outcome, this algorithm creates a ranker in the form of a decision tree. This tree can be used for classifying a new sample by beginning from the root of the tree and moving through its branches until a leaf node is reached (Kumar et al., 2016). In each non-leaf decision node, the outcome of characteristics is determined and used for testing in the node, whereby the attention is shifted toward the root below the selected tree (Chang et al., 2006).

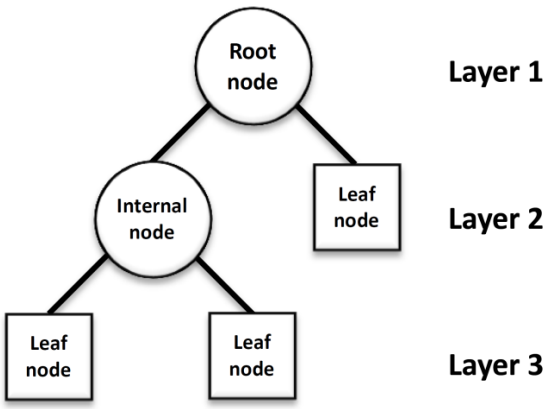

Figure 1. Classification of a new sample based on the decision tree model

The framework of the C4.5 algorithm is based on the method of Hunt CLS to create a decision tree from a set ‘T’ of the training samples. If the groups and classes are represented as C1, C2, ..., Ck, there will be three possibilities for the content of set ‘T’:

1. ‘T’ includes one sample or more, all belonging to a single group (class) Cj. Thus, the decision tree for T is a leaf identifying Cj group.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

2. ‘T’ does not contain any sample, and the decision tree is again a leaf. However, the group that corresponds to the leaf should be determined from other data of ‘T’, including the general majority group in ‘T’. The C4.5 algorithm is used as a criterion for most normal groups and indeed a specific group.

3. ‘T’ includes samples belonging to a combination of groups. In these conditions, the main idea is to refine ‘T’ in a subset of samples proceeding towards a single-group set of samples. It is selected based on a single attribute and a suitable test with interactional specific result(s) {O1, O2,..., On}. ‘T’ is categorized into trees T1, T2, ..., and Tn trees, where T1 includes all samples in ‘T’ with Oi results of the selected test. The decision tree for ‘T’ includes a decision node, identifying the necessary test and revealing a branch for every possible outcome (Campagni et al., 2014).

Generally, the C4.5 presents solutions for three types of tests:

1. A “standard” test based on a discretization attribute with an outcome or branch for any possible value of that attribute.

2. If attribute Y has continuous numerical values, a binary test can be defined with the results and incidents $Y \leq Z$ and $Y>Z$ by comparing its value against a ‘Z’ threshold value.

3. It is a more complex test based on a discrete activity in which the possible values are assigned to a certain valid number of groups with an outcome, an incident, and a branch for each group (Buldu et al., 2010).

2.3. Data envelopment analysis

Data envelopment analysis (DEA) is a method based on mathematical programming. It estimates the technical efficiency and inefficiencies of a unit. Without determining any assumption of the production function, this method estimates the production function or cost function through piecewise covering by solving mathematical models for a set of decision-maker units based on information related to the volume of real inputs and outputs of those units. Charns and Cooper (1978) in their famous paper introduced DEA as a mathematical programming model which determines the relative efficiency of a set of decision-maker units with multiple definite inputs and outputs of similar category. DEA compares the inputs and outputs with each other using observable data and by performing a series of optimizations to measure and assess the efficiency of each unit (Portela et al., 2004).

The tested decision-making units are independent units which employ inputs for producing similar outputs. The first prerequisite in selecting the assessed units is the homogeneity of inputs and outputs, since all of these units and the resultant envelopes are evaluated in one space. The boundary obtained is indeed the boundary of efficiency. The points on the boundary represent

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

efficiency. Other units that exist across the envelope area are inefficient and should fall on the boundary to be efficient (Wu et al., 2006).

In the relative measurement of units, Farel focused on the sum of units to create a virtual unit. The following relation was proposed as a common measurement tool for technical efficiency measurement.

Efficiency = weighted sum of outputs / weighted sum of inputs (1)

If the goal is to investigate the efficiency of n units with each containing m inputs and s outputs, the efficiency of the jth unit can be calculated as follows:

$$\text{Efficiency of the jth unit} = \frac{\sum_{r=1}^{s} U_r Y_{rj}}{\sum_{i=1}^{m} V_i X_{ij}} \quad (2)$$

To calculate the efficiency of every DMU (Decision Making Unit), the maximum efficiency index, i.e. the ratio of the sum of weighted outputs (virtual output) to the sum of weighted inputs (virtual input) should be obtained. Thus, for the evaluated unit, herein called unit zero, we will have:

$$MaxZo = \frac{\sum_{r=l}^{s} U_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \quad (3)$$

Subject to:

U r ≥ 0 r = 1,2,…,s

v i = ≥ 0 i = 1,2, …,

In the relation, Ur and Vis are the variables of the problem and weights, where the solution of the problem offers the most suitable value for the unit zero weights and thus measures its efficiency. In the model, the output coefficients are likely to be very large while the input coefficients are likely to be very small. Thus, to prevent such a problem, all efficiency ratios of units are considered as smaller than or equal to 1, which means that the model should be constrained. Within the constraints, any desired positive number such as k can be set instead of 1. Thus, the above programming can change according to the following:

$$Max\ Zo = \frac{\sum_{r=l}^{s} U_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \quad (4)$$

Subject to:

$$\frac{\sum_{r=l}^{s} U_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \quad (5)$$

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

$u_r \geq 0, v_i \geq 0$; r =1,2,…s; i=1,2,…,m; j=1,2,…,n

The model is called the CCR ratio model, where Zo indicates the maximum possible efficiency for the assessed unit (unit zero). To linearize the model, the CCR fractional programming method can be used (Wang et al., 2011).

3. The proposed approach

The present study conducts its proposed approach within six stages. The input is a database. That database is comprised of 4169 pieces that were randomly under the quality control process within three months. The anti-roll production line comprises nine working stations including cutting quality, machining quality, rolling quality, cooling quality, soldering quality, crack detection quality, framing quality, forming quality, and drilling quality. The quality of performance at each station is categorized either as high, medium, low, or very low. Also, the quality of the piece is classified into three groups, i.e. a high-quality piece means a piece with no defects, a medium-quality piece suggests a need for correction, and a low-quality piece indicates a wasted piece. Finally, the output of the proposed approach involves ranking the waste rules and determining the most important characteristics affecting breakdown. Table 2 shows the stages of the proposed approach.

Table 2. the procedure of the study

Input: database
Stage 1: Preparing the target data set
Stage 2: implementing data mining algorithm on the database
Stage 3: determining characteristics affecting breakdown by implementing the decision tree
Stage 4: determining rules in depths 1, 2, and 3
Stage 5: forming the data envelopment analysis to assess the rules
Stage 6: ranking the rules by data envelopment analysis model

Stage 1: preparing a target data set: this stage is done after determining related features and the aims of data mining, cleaning the data, and unifying the data. After examining each characteristic and giving it a numerical value, they are put into different classes.

Stage 2: implementing the data mining algorithm based on a 10-point validation database and by the Clementine Software.

Stage 3: determining the most important characteristic affecting breakdown by implementing the decision tree

Stage 4: determining rules at depths of 1, 2, and 3

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Stage 5: conducting the data envelopment analysis to assess the rules

Stage 6: ranking the rules by data envelopment analysis

4. The results of the proposed approach

4.1. Descriptive results

The descriptive results on each of the attributes can be seen in Tables 2 and 3. Also, regarding the database class, the descriptive information can be observed in Table 3.

Table 3. The qualitative status of piece production in each of the stations

Attribute	high	medium	low	very low
Cutting quality	2095	946	720	408
Machining quality	2135	1050	609	375
Rolling quality	1663	1286	730	490
Cooling quality	2374	1086	448	261
Soldering quality	2189	1062	502	416
Crack detection quality	1247	1284	849	789
Molding quality	1351	1064	1020	734
Forming quality	1524	1445	576	624
Drilling quality	1226	1656	732	555

According to the results in Table 3, the features of each working station were described and, as it is observed, in each production quality station, different pieces are produced, either in high, medium, low, or very low quality.

For instance, regarding the characteristics of cutting quality among 4169 sampled pieces, 408 pieces are in the 2003-2004 mm scope, suggesting a very low quality in this characteristic, 720 pieces are in the 2002.9-2003.5 mm scope, suggesting a low quality in this characteristic, 946 pieces are in the 2002.4-2002.8 mm scope, suggesting a medium quality in this characteristic, and 2095 pieces are in the 2002.1-2002.3 mm scope, suggesting a high-quality for this characteristic.

Table 4. Percentage of quality of piece production in each of the stations

Attribute	High	Medium	Low	Very low
Cutting quality	50	23	17	10
Machining quality	51	25	15	9

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Rolling quality	40	31	18	12
Cooling quality	57	26	11	6
Soldering quality	53	25	12	10
Crack detection quality	30	31	20	19
Molding quality	32	26	24	18
Forming quality	37	35	14	15
Drilling quality	29	40	18	13

According to the results in Table 4, the quality (%) of each feature is shown, resulting from each working station. As observed, the highest percentage of very low quality is related to the crack detection quality, whereas the lowest percentage of the same is related to the cooling quality.

Table 5. The status of class (final product quality) of the database

	Frequency	Percent
High	3496	83.85704
Low	241	5.780763
Medium	432	10.3622
Total	4169	100

According to the results in Table 5, it is suggested that approximately 84% of the products have a proper quality while 10% showed a medium quality. Nearly 6% of the products had significant problems. Thus, the present study aimed to predict and improve efficiency in the main working station.

4.2. Data mining results

By implementing a 10-point validation scale, the accuracy of the C5 algorithm was determined (Table 4).

Table 6. The accuracy of the C5 algorithm

	No.	percent
Correct	3.921	94.05
Wrong	248	5.95
Total	4.169	100

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

As shown in Table 6, the C5 tree algorithm has an appropriate accuracy in predicting the quality of the pieces.

4.3. Determining the most important attributes

The most important attributes affecting the accuracy of prediction in the database class can be observed in Figure 3. The most important features in the accuracy of quality prediction for the pieces are cooling, soldering, and forming. When the most important features were determined, the rules of the decision tree were extracted.

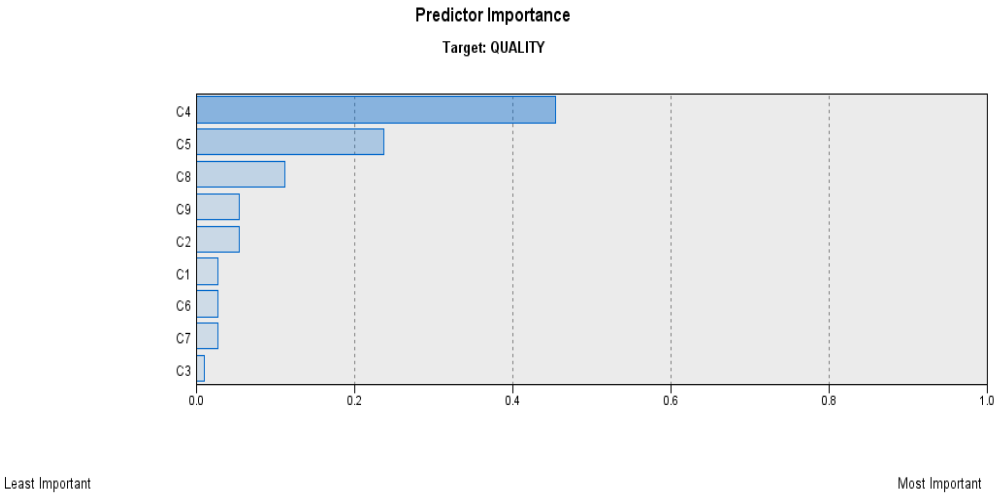

Figure 3. The most important attributes affecting the quality

According to Figure 3, the most important features in determining the accuracy of quality prediction for the pieces were cooling, soldering, and forming, respectively.
Figure 4 shows the results of the decision tree.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

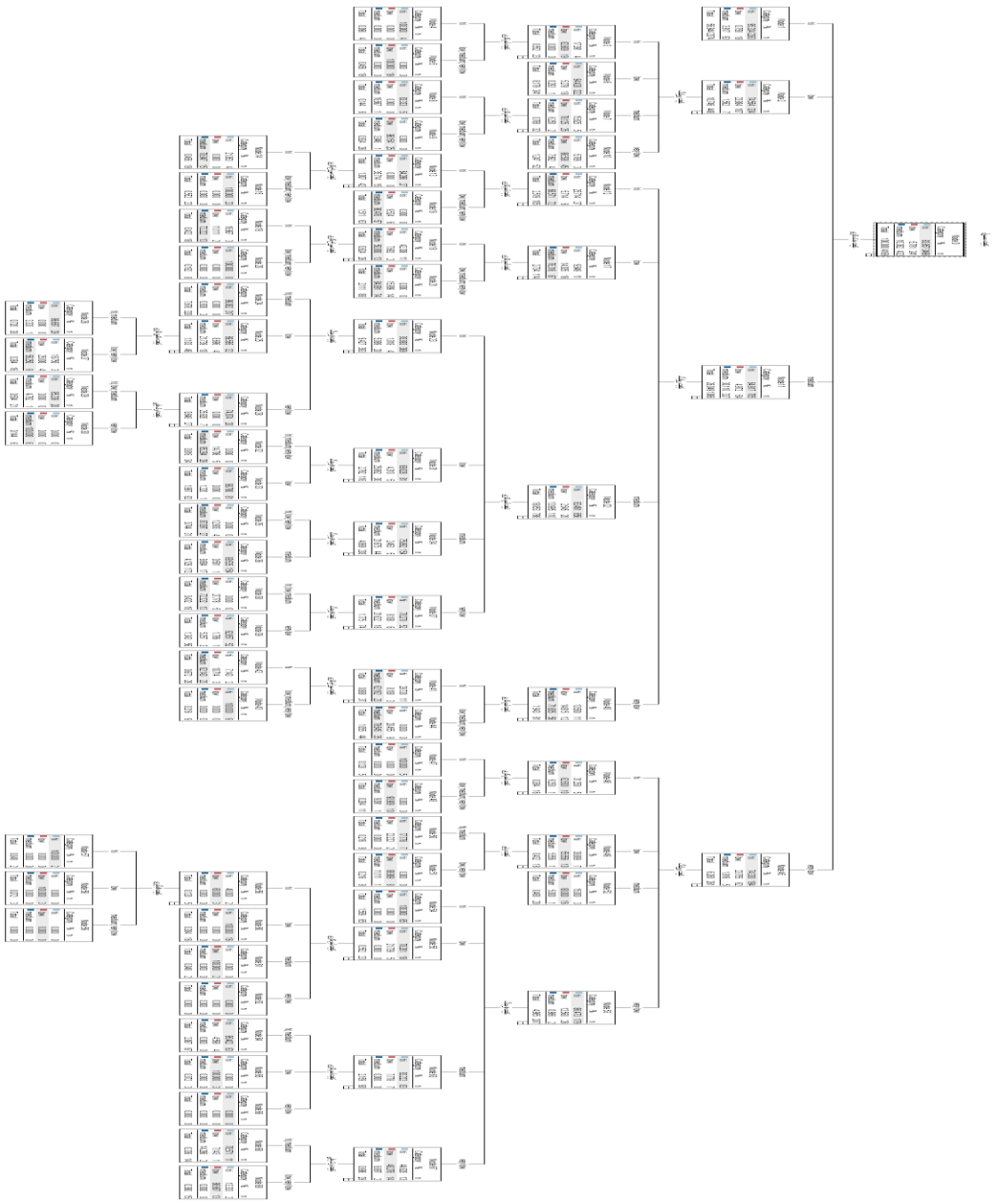

Figure 4. C5 decision tree

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Tables 7 to 9 show the rules in depths of 1 to 3.

Table 7. The rules with depth 1

the condition of the new node		Node No.	No. of defects	No. of reworking	Probability of defect	Probability of reworking
Cooling quality	high	1	18	93	1	4
	Low	2	107	7	24	2
	Medium	11	54	327	5	30
	Very low	45	62	5	24	2

According to Table 7, node 2 must be taken into account, because there are 107 defective pieces in this node which constitute nearly 40% of the defective pieces.

Table 8. The rules in depth 2

condition	Condition of new node		Node No.	No. of defects	No. of reworking	Probability of defect	Probability of reworking
Low cooling quality	Rolling quality	High	3	19	0	83	0
		Low	6	18	1	5	0.3
		Medium	7	25	2	78	6
		Very low	10	45	4	87	6
Medium cooling quality	Rolling quality	High	12	6	72	6	69
		Low	17	16	87	14	76
		Medium	22	20	110	3	14
		Very low	40	12	58	15	72
Very low cooling quality	Rolling quality	High	46	10	1	63	6
		Low	49	10	1	56	6
		Medium	52	16	1	80	5
		Very low	53	26	2	13	1

Table 8 shows the examination of the rules in depth 2. Accordingly, 12 rules were determined.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Table 9. The rules in depth 3

Condition	New node condition		Node No.	No. of defects	No. of reworking	Defect probability	Reworking probability
Low cooling quality and medium rolling quality	Soldering quality	High	8	0	1	0	17
		Not high	9	25	1	96	4
Low cooling quality and high rolling quality	Soldering quality	High	4	0	0	0	0
		Not high	5	19	0	100	0
Medium cooling quality and high rolling quality	Soldering quality	High	13	0	15	0	36
		Not high	16	6	57	10	90
Medium cooling quality and low rolling quality	Soldering quality	High	18	2	13	8	50
		Not high	21	14	74	16	84
Medium cooling quality and medium rolling quality	Soldering quality	High	23	4	1	20	5
		Low	31	5	30	4	26
		Medium	34	5	44	2	21
		High low	37	6	8	16	22
Medium cooling quality and very low rolling quality	Soldering quality	High	41	3	23	8	62
		Not high	44	9	20	35	80
Very low cooling quality and high rolling quality	Soldering quality	High	47	0	0	0	0
		Not high	48	10	1	90	9
Very low cooling quality and low rolling quality	Cutting quality	High and medium	50	2	0	22	0
		Low and very low	51	8	1	89	11

Table 9 shows the examination of the rules in depth 3. Accordingly, 18 rules were determined, and then, the rules were assessed by conducting a data envelopment analysis in two phases. In general, the output-input pattern in the model was outlined according to the following:

The input is the constant value of 1.
The output is the value that contains four different variables, including the number of defects, the number of reworking, the probability of the defects, and the probability of reworking. In the first phase, the efficiency of each rule is determined in the nodes at a similar depth. In the second phase, the efficiency of each rule is specified across all depths (Tables 10 and 11).

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

Table 10. The efficiency and ranks of rules in each depth

Node No.	Efficiency in similar depth	Rank in a similar depth
1	0.309	4
2	1	1
11	1	1
45	1	1
3	0.954	5
6	0.4	12
7	0.903	8
10	1	1
12	0.908	7
17	1	1
22	1	1
40	0.966	4
46	0.743	9
49	0.668	10
52	0.92	6
53	0.578	11
8	0.259	16
9	1	1
4	0	17
5	1	1
13	0.621	10
16	1	1
18	0.754	9
21	1	1
23	0.35	14
31	0.591	11
34	0.46	13
37	0.52	12
41	0.91	8
44	1	1
47	0	17
48	1	1
50	0.3	15

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

51	1	1

Table 10 shows the efficiency of the rules determined in Tables 7 to 9. Accordingly, when the tree depth is compared in homogeneous nodes – at the same tree depth – most efficiencies equal to 1, due to the fewer number of nodes available for comparison. In the other depths, however, the efficiency dropped significantly from 1.

Table 11. Efficiency and the general rank of rules

Node NO.	General efficiency	General rank
1	0.309	26
2	1	1
11	1	1
45	0.658	19
3	0.845	14
6	0.175	32
7	0.85	13
10	1	1
8	0.189	31
9	1	1
12	0.828	16
17	0.948	11
22	0.393	23
40	0.863	12
46	0.679	18
49	0.609	20
52	0.841	15

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

53	0.297	28
4	0	33
5	1	1
13	0.4	22
16	1	1
18	0.566	21
21	1	1
23	0.241	29
31	0.323	25
34	0.3	27
37	0.349	24
41	0.694	17
44	1	1
47	0	33
48	0.973	10
50	0.22	30

Table 11 shows the efficiency of all rules determined in this context. Accordingly, at all depths, there are rules with an efficiency of 1, suggesting that the rules were homogeneous and took all nodes into account.

The data were analyzed using two software applications. In data mining and the implementation of the decision tree algorithms, the Clementine Software was applied as a specialized data mining application. Also, regarding the data available for the analysis, the DEAP software offered a specialized application for running the data envelopment analysis.

5. Discussion and interpretation of the findings

In this research, by combining data mining and data envelopment analysis, the product quality was predicted and the most important rules leading to low-quality at low and medium levels were examined. Based on the results obtained from the C5 tree, 34 rules were evaluated at the 1st to the

Power System Technology
ISSN:1000-3673
Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

3rd levels of the decision tree. Based on the ranking obtained at each depth, it can be stated that at depth 1, except for node 1, all nodes were regarded as important factors for the prediction of low product quality. Considering depth 2, nodes 10, 17, and 22 had a similar status, and eventually, at depth 3, nodes 9, 5, 44, 48, and 51 were important for preventing low product quality. Based on the general ranking, it can be deduced that nodes 2, 11, 20, 5, 16, 5, 21, and 44 were implicated in low and medium quality. In other words, to prevent defects in the product, attention should be paid to these nodes so that negative attributes do not occur in these nodes. Future research is suggested to deal with factors such as the cost and implementation of node assessment and determining the executive costs of preventing possible defects.

Based on the results obtained in the present study, the proposed approach can be used to offer functional solutions to increase the quality of the pieces and minimize waste in production lines. A close look at Figure 3 shows that the most important features affecting breakdown are determined by the c5 tree. In rules with a high potential in piece breakdown (rules 2, 11, 10, 9, 5, 16, 18, 44), most of the characteristics such as cooling, soldering, and forming are present either alone or with another characteristic. Thus, it is recommended that at least one working station among these three working stations be improved. In addition, a highly efficient approach would be to build a smart system to prevent the breakdown before production, thereby enabling the rules to run at depths of 3 or 4. This might bring about some challenges for the manufacturer but will eventually decrease the waste of pieces.

6. Validation of the model

The model was validated by implementing changes in the anti-roll bar and soldering priorities from April to September 2021. Table 12 shows the results of changes in quality.

Table 12. Results of production from April to September 2021

	The year 2019		**April to September 2021**	
	Number	Percentage	Number	Percentage
High	3496	83.85704	8421	90%
Low	241	5.780763	652.47	7%
Medium	432	10.3622	247.53	3%
Total	4169	100	**9321**	100

Table 11 shows that as the conditions were adequately sustained and the quality of input materials remained unchanged in 2021, the quality improved by 7%, suggesting that the model has an appropriate validity.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

7. Conclusion

While quality standards were determined in previous studies on enhanced manufacturing, the use of decision tree algorithms via the mechanism of these rules have not been explored in detail. In other words, optimizing the production lines via product quality optimization is not possible without considering the effective variables on decision-making, such as the percentage and number of each breakdown, given the performance of each working station. These issues were not covered considerably in previous studies. Thus, the priority of rules were specified via an appropriate planning model in the current research after extracting the rules. Then, the optimization of production was conducted accordingly. This study and its use of the decision tree as an input for the data envelopment analysis can be regarded as a new approach. Manufacturers can better achieve optimization in industrial automation by repeating the approach of the current study while reducing waste and increasing product quality. Future research can consider taking action toward mechanized solutions by means of the decision tree. The working stations can be reevaluated by techniques such as linear programming while taking into account the implementation of costs and challenges. The rules can be ranked again by taking into account cost-related parameters of breakdown and entering them into the data envelopment analysis.

References

[1] P. Podrzˇaj, S. Simoncˇicˇ, Resistance spot welding control based on fuzzy logic, Int. J. Adv. Manuf. Technol. 52 (9–12) (2011) 959–967.

[2] Andersen,K., Cook,G.E., Ramaswamy,K., Artificial Neural Networks Applied to Arc Welding Process Modeling and Control, IEEE Transaction on Industrial Application, 2011, Volume 26, pp. 824-830.

[3] Bouslah, B., Gharbi, A., & Pellerin, R. (2016). Integrated production, sampling quality control and maintenance of deteriorating production systems with AOQL constraint. *Omega*, *61*(4), 110–126.

[4] Devi, S. JothI, S, DevI. A, Data Mining Case Study for Water Quality Prediction using R Tool, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2018, Volume 3, 262-269.

[5] Loukas, K. Quality improvement calls data mining: the case of the seven new quality tools, Benchmarking: An International Journal, 2018, Volume, 47-75.

[6] Bordin, CH., Anuta, H., Crossland, A., Gutierrez, L., Dent, Ch and Vigo, D. (2017). A linear programming approach for battery degradation analysis and optimization in offgrid power systes with solar energy integration. Renewable Energy, Elsevier, vol. 101©, pages 417-430.

[7] Hang Yan, Nan Yang, Yi Peng, Yitian Ren, (2020), Data mining in the construction industry: Present status, opportunities, and future trends, Automation in Construction, Volume 119, 113-124.

Power System Technology
ISSN:1000-3673

Received: 04-02-2024 Revised: 08-04-2024 Accepted: 27-05-2024

[8] Baysal,M., Erdem,Gu¨nay,M., Yıldırım,R., Decision tree analysis of past publications on catalytic steam reforming to develop heuristics for high performance: A statistical review, International Journal O F Hydrogen Energy, 2017, Volume 4 2 , 2 4 3-2 5 4.

[9] Alves,M., Ramirez,M., Guimaraes,F., Escobar,A, Linear Programming And Genetic Algorithm For Generation Maintenance Scheduling And Hydrothermal Dispatch Considering Uncertainties In Multicriteria Decision Making, XLIX Simpósio Brasileiro de Pesquisa Operacional, Blumenau-SC, 2017, Volume 27 a 30,1-12.

[10] Kamikawaji,Y., Matsuyama,H., Fukui,K., Hosoda,S., Ono,S., Decision Tree-based Feature Function Design in Conditional Random Field Applied to Error Detection of Ocean Observation Data, Computational Intelligence (SSCI), 2016, Volume 4, 1-8.

[11] Tzu-Liang Tseng,B., Kalyan Reddy,A., Zhonghua,H., Yongjin,K., E-quality control: A support vector machines approach, Journal of Computational Design and Engineering, 2016, Volume 3, 91–101.

[12] Huang,W., Zhang,Q., Xu,W., Fu,H., Wang,M.,Liang,X,A Novel Trigger Model for Sales Prediction with Data Mining Techniques,Data Science Journal. 2016, Volume 14, p.15. DOI: http://doi.org/10.5334/dsj.

[13] Kim,S., Jipitaklert,W., Park,S., Hwang,S., Data mining model-based control charts for multivariate and autocorrelated processes,Expert Systems with Applications, 2014, Volume **39, Issue 2, Pages 2073–2081.**

[14] Kumar,D., Rahman,Z., Chan,F, A fuzzy AHP and fuzzy multi-objective linear programming model for order allocation in a sustainable supply chain: A case study, International Journal of Computer Integrated Manufacturing, 2016, Volume 30, Issue 6, 535-551.

[15] Chang, K.-M., Beck, J., Mostow, J., Corbett, A., 2006. A bayes net toolkit for student modeling in intelligent tutoring systems. In: Paper presented at the Intelligent tutoring systems. European Journal Of Operational Research. (2), 429-444.

[16] Campagni, R., Merlini, D., Verri, M.C., 2014. An Analysis of Courses Evaluation Through Clustering. In: Paper presented at the International Conference on Computer Supported Education

[17] Buldu, A., Üçgün, K., 2010. Data mining application on students' data. Proc. Social Behav. Sci. 2 (2), 5251–5259.

[18] Portela, M. S., Thanassoulis, E., & Simpson, G. (2004). Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society*, *55*(10), 1111-1121.

[19] Wu, D. D., Yang, Z., & Liang, L. (2006). Efficiency analysis of cross-region bank branches using fuzzy data envelopment analysis. *Applied Mathematics and Computation*, *181*(1), 271-281.

[20] Wang, Y. M., & Chin, K. S. (2011). Fuzzy data envelopment analysis: A fuzzy expected value approach. *Expert Systems with Applications*, *38*(9), 11678-11685.