PROBLEMY MECHATRONIKI
UZBROJENIE, LOTNICTWO, INŻYNIERIA BEZPIECZEŃSTWA

ISSN 2081-5891 **12, 2 (44), 2021, 39-54**

PROBLEMS OF MECHATRONICS
ARMAMENT, AVIATION, SAFETY ENGINEERING

Prediction of Post-Diagnostic Decisions for Tested Hand Grenades' Fuzes Using Decision Trees

Dariusz AMPUŁA

Military Institute of Armament Technology
7 Prymasa St. Wyszyńskiego Str., 05-220 Zielonka, Poland
Author's e-mail address and ORCID:
ampulad@witu.mil.pl; https://orcid.org/0000-0002-9036-9498

Received by the editorial staff on 3 December 2020
The reviewed and verified version was received on 1 June 2021

DOI 10.5604/01.3001.0014.9332

Abstract. The article presents a brief history of creation of decision trees and defines the purpose of the undertaken works. The process of building a classification tree, according to the CHAID method, is shown paying particular attention to the disadvantages, advantages, and characteristics features of this method, as well as to the formal requirements that are necessary to build this model. The tree's building method for UZRGM (Universal Modernised Fuze of Hand Grenades) fuzes was characterized, specifying the features of the tested hand grenade fuzes and the predictors used that are necessary to create the correct tree model. A classification tree was built basing on the test results, assuming the accepted post-diagnostic decision as a qualitative dependent variable. A schema of the designed tree for the first diagnostic tests, its full structure and the size of individual classes of the node are shown. The matrix of incorrect classifications was determined, which determines the accuracy of incorrect predictions, i.e., correctness of the performed classification. A sheet with risk assessment and standard error for the learning sample and the v-fold cross-check were presented.

On the selected examples, the quality of the resulting predictive model was assessed by means of a graph of the cumulative value of the lift coefficient and the "ROC" curve.
Keywords: mechanical engineering, decision trees, branch, leaf, node, feature

1. INTRODUCTION

One of the methods of constructing classification [1] rules is the method based on classification trees and it consists in gradual division of a set of objects into subsets until their homogeneity is achieved due to belonging to specific classes.

Classification trees were created in the early eighties as a result of searching for methods that imitate learning and problems' solving by people. The main ideas, however, come from the sixties, when the idea of using a tree structure (they were called decision trees) to represent the process of creating concepts arose. Then, E.B. Hunt, J. Marin, and P.J. Stone built the CLS (Concept Learning System) algorithm (Hunt et al. 1966). The CLS algorithm has become an inspiration for conducting further research in this direction, not only in psychology.

The most important stage in the development of these methods was the appearance of the Quinlan ID3 algorithm (1983), whose successful practical applications highlighted classification trees as a convenient data classification tool. At the same time, on the basis of statistics, a search was made for reference classification methods that would be less demanding than discriminatory functions. The result was the C&RT (CART) algorithm of Breiman, Friedman, Olshen, and Stone (1984).

Another method of building classification trees is the CHAID (Chi-square Automatic Interaction Detector) method, proposed by Kass (1980) according to Ripley (1996), which is the successor of the THAID algorithm created by Morgan and Messenger (1973).

The CHAID method [2] is another method of data analysis based on the AID interaction detection method, which was introduced by Sonquist and Morgan (1964). This method allows you to divide the set of cases into comprehensive and mutually disjoint subsets that best describe the dependent variable. Other features characterising this method are the way of splitting nodes (using the chi-square independence test) and the possibility of building non-binary trees, i.e., it does not build only binary trees but also the trees in which more than two branches can emerge from the nodes.

A classification tree can be defined as a tree representing the process of dividing a set of objects into homogeneous classes. Its internal nodes describe how to make this division (based on the object feature values), and the leaves correspond to the classes to which the objects belong. In turn, the edges of the tree represent the values of the features on the basis of which the division was made.

Classification trees have a very useful feature, namely they can be used to classify a new object, but it is not required to know all the features of this new object. All methods that create classification trees have a very similar structure. We can say that they are based on solutions that included the first three algorithms: CLS, ID3, and C&RT.

The differences concern, among others, the form of the function assessing the quality of the division and the method of classification of objects with specific missing values of features.

The purpose of the work was to design and build a decision tree based on the results of laboratory tests of hand grenade fuzes. To achieve this target, UZRGM fuzes were taken, whose data base is the largest among all fuzes tested, which gives the probability of developing a decision tree with a high quality level of its work. The classification tree, built using the CHAID method, shows us the possibility of using the theory of these decision trees to support making post-diagnostic decisions for the tested hand grenade fuzes. The tree, designed in this work, has concerned the evaluation module of the tested fuzes for the first laboratory diagnostic tests. The secondary target of the designed tree was the prediction of post-diagnostic decisions for new predictor values obtained during testing new lots of UZRGM fuzes for hand grenades.

2. THE METHOD OF BUILDING DECISION TREES

Classification trees [1] are created by recursively dividing a set into subsets until they are homogeneous due to the objects' affiliation to specific classes. The point is that such a tree should be as small as possible (it has a minimum number of nodes), i.e., that the classification rules obtained are as simple as possible.

The effectiveness of the classification tree creation algorithm depends on the choice of the way of dividing the sets of objects in the tree nodes, i.e., individual features or their linear combination. This selection is based on a certain measure of the quality of the division. In practice, homogeneity measures or measures of differentiation of subsets, obtained as a result of division, are used for this purpose.

After the tree is constructed, a decision should be made when to stop subdividing the subsets. The idea is to obtain a tree with a minimum number of nodes, without reducing the "quality" of the classification rules of the analysed objects.

In the CHAID method, at each stage of the tree division, a contingency table is created in which the dependent and independent variables are combined. For example, if the dependent variable has $d \geq 2$ categories and the predictor $c \geq 2$ categories, it is aimed to reduce the resulting contingency table with the dimensions d x c to the more significant with the dimensions d x j, by combining the predictor categories in a permitted manner.

If n predictors were included in the analysis, then n reduced contingency tables are obtained. For each of them, the Pearson chi-square test of independence is used and the test probability value p (p value) is calculated. Then, the adjusted p value is estimated, which is the product of the p level and the Bonferroni multiplier. This multiplier is calculated differently for each type of a predictor.

One of the advantages of the CHAID method [2] is the ability to build trees with any number of branches. However, research practice shows that trees of this type are very often binary models.

The characteristic features of the CHAID method described in the article are the way of dividing nodes (using the chi-square independence test) and the possibility of building non-binary trees (with any number of branches). In addition to the ordinal and nominal predictors used in this method, the so-called variable predictors are introduced. The floating predictor is an independent variable at the ordinal measurement level, which has the so-called floating category. This is a category that indicates an unknown item on the scale or no data that can be combined with any rating on the scale in any order.

The interaction detection method used, basing on the chi-square classification, has several limitations. The disadvantage was the possibility of creation only discriminatory models. The dependent variable had to be on the nominal level of its measurement. Some researchers saw in this the benefits associated with the discretisation of metric variables with an asymmetrical or bimodal distribution.

The CHAID method should not be used to analyse small sets of observations, i.e., for the sets smaller than 1000 cases, and if cross-evaluation is enabled this threshold increases to 2000 observations. In addition, another disadvantage is the sensitivity of the Bonferroni multiplier to the number of predictors and the number of observations. This multiplier is used in calculating the significance of the node division. Too many predictors and too few observations can lead to an incorrect estimation of its value. In the case of UZRGM hand grenade fuzes, analysed in the article, this is not the case. We have over 2000 observations here and only six predictors.

3. METHODOLOGY OF BUILDING A DECISION TREE FOR FUZES OF UZRGM TYPE

When designing the decision tree for the first diagnostic laboratory tests, the results of UZRGM fuzes were prepared [3, 4]. This type of fuzes are used in hand grenades: F-1, RG-42, RGO-88, RGZ-89, and CGR-42A.

The results of the so-called scientific-research showed that they are not authoritative to other test results. The diagnostic tests, carried out for the Ministry of the Interior, were not analysed.

Only the tests in which the type of test specified in the test methodology [5] was equal to one of the test samples stored in the warehouses of the economic branches of the Polish Army were taken for the analysis, which means that only the examined fuze lots stored in the storage subset specified as "K" were taken into account. All these limitations were aimed at creating a homogeneous data set that could be analysed by the designed decision trees.

The UZRGM fuze is a time fuze which, in its design, has a certain delay in operation (the fuze's delay time). It consists, according to the description [6], of three basic assemblies: impact device, safety mechanism, and ignition device. The delay of this fuze ranges from 3.2 seconds to 4 seconds and it is the basic parameter checked during the laboratory diagnostic test. Another element tested is the spring testing, which is structurally used inside the fuze. The correct operation of the ignition primer and excitation primer, the operation of the fire chain are also checked, as well as the corrosion of individual parts and assemblies of the fuze, the state of the fuze's protection and the correct operation of the fuze needle.

All properties (features) of the UZRGM fuze, according to the test methodology, were divided into four classes of validity (inconsistencies): A, B, C, and E. Depending on the number of detected inconsistencies, in individual importance classes during laboratory tests, it was obtained a specific post-diagnostic decision, according to the module's evaluation.

In our case, for the designed decision tree and the analysed UZRGM fuzes, as a result of the first laboratory diagnostic tests, 7 different data (predictors) of the tested features were accepted, which were the information obtained after the diagnostic tests, namely the predictors: total number of inconsistent fuzes (N), number of inconsistencies in the importance class A (LA), number of inconsistencies in the importance class B (LB), number of inconsistent fuzes in the importance class B (NB), number of inconsistencies in the importance class C (LC), number of inconsistent fuzes in the importance class C (NC), and a number of inconsistencies in the importance class E (LE).

Therefore, in our case, the results of all the tested features of a given fuzes' lot were the values of the predictors. These parameters were written in a numerical form, i.e., if during the diagnostic test no inconsistencies of a given class were found, then the value zero was provided.

However, if inconsistencies were found in the tests, then a specific number of these inconsistencies was given. The searched value was the specific post-diagnostic decision obtained in accordance with the test methodology [5].

In our designed decision tree, we will deal with a classification tree due to the fact that it is possible to obtain several different post-diagnostic decisions, depending on the number of inconsistencies received during laboratory tests. According to the evaluation module in the test methodology, as a result of the conducted first laboratory tests, six different post-diagnostic decisions can be obtained.

During building our classification tree, additional auxiliary parameters were accepted, whose task was to design the best classification tree for the test results of the analysed fuzes. The values of these parameters have been entered into the software [7].

To sum up, the subject of classification during the design and building our tree was a set of data obtained during the first laboratory tests of hand grenade fuzes of UZRGM type. Each tested lot of fuzes was characterized by the obtained results of the tested features of these fuzes, recorded in a numerical form and entered into the database. A classification tree, built using the CHAID algorithm, has been designed. Trees of this type are designed for observations whose number is at least 1000, then the tree with the best predictive parameters is created.

4. THE RESULTS OF BUILDING DECISION TREE BY THE CHAID METHOD

For building our decision tree, the CHAID algorithm was accepted, whose operation consists in dividing the set of the analysed cases into comprehensive and mutually disjoint subsets, the best describing a dependent variable. In our case, a qualitative dependent variable, designated as "DEC" was accepted, which means the post-diagnostic decision obtained after the first laboratory tests. The above decision may take the form of six different decisions: "B5", "B3", "BP", "Z", "PS", and "W". The exact description of possible diagnostic decisions was presented in the test methodology mentioned earlier. All records in the UZRGM fuze test database, which was used when designing this tree, were prepared according to the same key, so that they formed a certain homogeneity and integrity. During classification process, there were accepted the so-called incorrect classification costs at the level "equal". With this option, you can give more weight to accurate prediction (classification) for the selected classes than for others ones.

The accepted value "equal" means giving one to all elements of the matrix of incorrect classification costs, except for the main diagonal of this matrix.

The next element to be defined is the stop criterion, which allows you to control the growth of the being built tree. Tree's size is an important computational issue, too large trees are difficult to interpret. In the stop criterion, the minimum number of nodes was assumed at the level of "50", which means the value used to control the end of divisions. The maximum number of nodes, at the level "1000", was also marked, which means the number determining at what number of nodes in the tree the algorithm will stop working. The values of the probability level "p" for dividing and the probability level "p" for joining were left according to the software suggestions, i.e., at the level of 0.05, because chi-square statistics are used here and it can be assumed that these default probability settings may remain unchanged.

The next step in tree design was to specify the validation method. A v-fold cross check was determined in which the number "10" was entered, which specifies the number of subsets to be used to assess the cost of the test for each tree created in the tree series being created.

The initial value of the random number generator at level "1" was accepted, which tells us about the value of the kernel used in the process of random grouping of data in 10 subsets defined by us.

In the last step, there has been accepted the so-called Bonferroni's correction, which is used to "hinder" recognition as a statistically significant result of a single test, when testing repeatedly based on the same data. This correction is taken into account when calculating the "p" value, i.e., the test probability level.

As a result of completing the process of building our tree's model, the tree with the best parameters has been selected, the scheme of which is shown in Fig. 1.

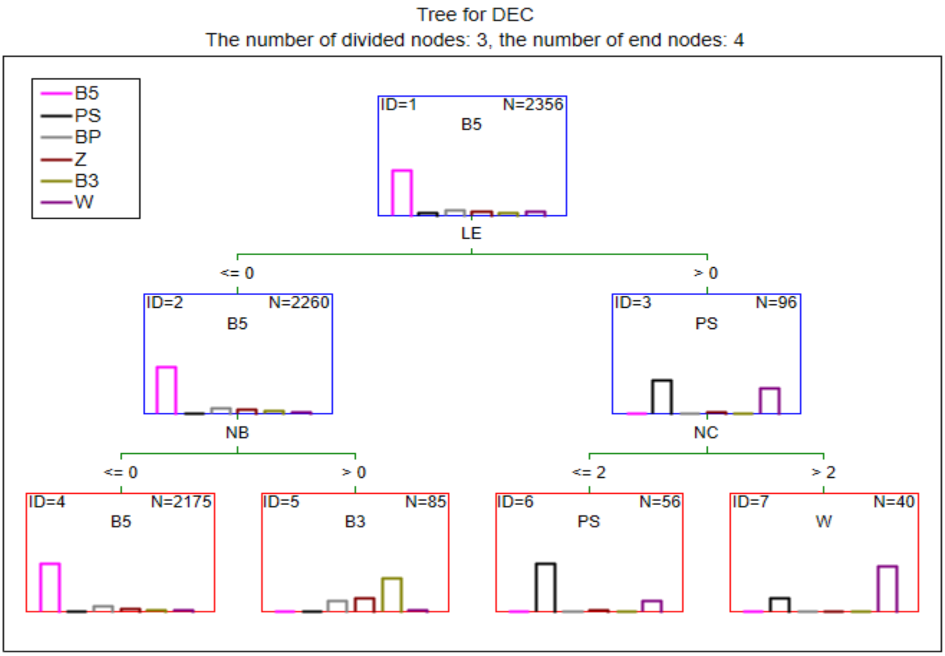

Fig. 1. The schema of the tree for the first laboratory tests

It has three divide nodes and four end nodes (leaves). The selected end tree was built basing on the given predictor sizes. Each node contains the node ID, node size, selected dependent variable category, and a histogram of dependent variables selected for the given node.

The exact numerical structure of the designed tree is shown in Fig. 2.

This table shows description of the data in all nodes of the built tree, i.e., the size of a given node, the number of individual accepted classes of the node, the selected class in a given node as well as the criterion for descendants and the number of descendants to which given cases may be sent, depending on which division criterion they meet.

The structure of the tree (UZRGM RB=1)
Dependent variable: DEC
Options: Quality dependent

	Number of nodes	Node size	N class B5	N class PS	N class BP	N class Z	N class B3	N class W	Chosen class	Divide variable	Criterion for child 1	Criterion for child 2	Child node 1	Child node 2
1	2	2356	1757	56	206	122	104	111	B5	LE	x <= 0	x > 0	2	3
2	2	2260	1757	2	206	121	104	70	B5	NB	x <= 0	x > 0	4	5
4		2175	1757	2	190	101	57	68	B5					
5		85	0	0	16	20	47	2	B3					
3	2	96	0	54	0	1	0	41	PS	NC	x <= 2	x > 2	6	7
6		56	0	45	0	1	0	10	PS					
7		40	0	9	0	0	0	31	W					

Fig. 2. The structure of the tree for the first laboratory tests

Matrix of classification (UZRGM RB=1)
Dependent variable: DEC
Options: Quality dependent, Attempt to analyze

	Observed	Predicted B5	Predicted PS	Predicted BP	Predicted Z	Predicted B3	Predicted W	Together in the line
Number	B5	1757						1757
% from column		80.78%	0.00%			0.00%	0.00%	
% from line		100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
% from total		74.58%	0.00%	0.00%	0.00%	0.00%	0.00%	74.58%
Number	PS	2	45				9	56
% from column		.9%	80.36%			0.00%	22.50%	
% from line		3.57%	80.36%	0.00%	0.00%	0.00%	16.07%	
% from total		.8%	1.91%	0.00%	0.00%	0.00%	0.38%	2.38%
Number	BP	190				16		206
% from column		8.74%	0.00%			18.82%	0.00%	
% from line		92.23%	0.00%	0.00%	0.00%	7.77%	0.00%	
% from total		8.06%	0.00%	0.00%	0.00%	0.68%	0.00%	8.74%
Number	Z	101	1			20		122
% from column		4.64%	1.79%			23.53%	0.00%	
% from line		82.79%	0.82%	0.00%	0.00%	16.39%	0.00%	
% from total		4.29%	.4%	0.00%	0.00%	0.85%	0.00%	5.18%
Number	B3	57				47		104
% from column		2.62%	0.00%			55.29%	0.00%	
% from line		54.81%	0.00%	0.00%	0.00%	45.19%	0.00%	
% from total		2.42%	0.00%	0.00%	0.00%	1.99%	0.00%	4.41%
Number	W	68	10			2	31	111
% from column		3.13%	17.86%			2.35%	77.50%	
% from line		61.26%	9.01%	0.00%	0.00%	1.80%	27.93%	
% from total		2.89%	0.42%	0.00%	0.00%	.8%	1.32%	4.71%
Number	Total groups	2175	56			85	40	2356
% together		92.32%	2.38%	0.00%	0.00%	3.61%	1.70%	

Fig. 3. The matrix of incorrect classifications for the first laboratory tests

The next step in the analysis of the built tree model was to assess the accuracy of the prediction. The simplest tool to assess the correctness of classification [2] is the resulting matrix of incorrect classifications (Fig. 3). This matrix compares the observed classes and the predicted classes. In our case, we received automatically calculated prediction errors, which are located in the "% from line" for individual predicted post-diagnostic decisions.

For example, for the observed "Z" class, the predictive error for the predicted "PS" class is 0.82%, while for the observed "W" class, the predicted error for the predicted "B3" class is 1.8%. Of course, the usefulness of a given built tree's model is determined not only by the accuracy of the prediction of the entire solution, but also by the accuracy of the prediction of individual model classes.

The calculated prediction errors, shown in Fig. 3, arising in a matrix of incorrect classifications, can also be presented in the form of a three-dimensional histogram, whose form is presented in Fig. 4, thanks to the software [7].

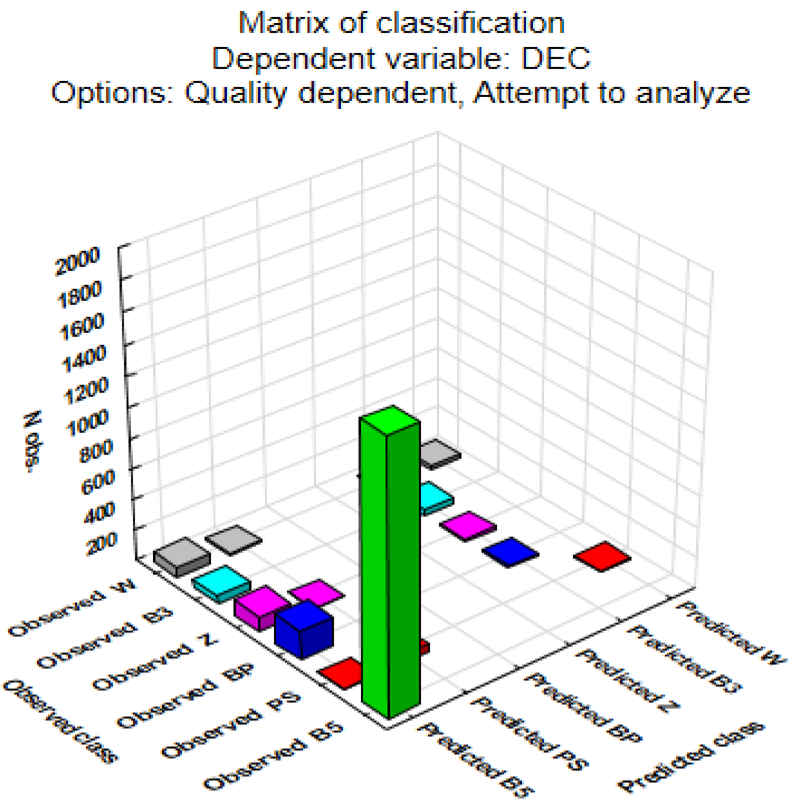

Fig. 4. Histogram of the predicted frequencies relative to the observed ones

Next, we determine the sheet with risk evaluation (Fig. 5) for our learning test and the risk of a v-fold cross-check that was previously included in the analysis. The risk is calculated as the fraction of cases wrongly qualified by the tree at the equal costs of incorrect classification we have marked.

	Risk evaluation (UZRGM RB=1) Dependent variable: DEC Options: Quality dependent	
	Risks Evaluation	Standard error
Learning	0.202037	0.008272
V-fold	0.157537	0.007507

Fig. 5. Risk evaluation sheet for the built tree

The next step is the ability to create a sheet of values observed and predicted by our built tree. In this sheet, the final node, observed class, predicted class, classification probabilities for each category, and posteriori probabilities for each class will also be given in the results for each case. A fragment of such a data sheet is shown in Fig. 6, which shows that there are discrepancies (in red) between the observed values and the predicted values, which should be verified by re-checking the correctness taken of the post-diagnostic decision. These discrepancies can be a mistake made by people making post-diagnostic decisions.

	Prediction (UZRGM RB=1) Dependent varaible: DEC Options: Quality dependent, Attempt to analyze								
	Observed value	Predicted value	Probab. B5	Probab. PS	Probab. BP	Probab. Z	Probab. B3	Probab. W	End node
1191	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1192	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1193	Z	B3	0.000000	0.000000	0.188235	0.235294	0.552941	0.023529	5
1194	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1195	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1196	W	PS	0.000000	0.803571	0.000000	0.017857	0.000000	0.178571	6
1197	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1198	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1199	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1200	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1201	W	W	0.000000	0.225000	0.000000	0.000000	0.000000	0.775000	7
1202	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1203	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1204	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4
1205	B5	B5	0.807816	0.000920	0.087356	0.046437	0.026207	0.031264	4

Fig. 6 Predicted value sheet by the built tree

In addition, we can create a graph for each node in the tree, which shows a diagram of observation values for the selected node for the given predictors. An example of such a graph for node number 1 is shown in Fig. 7. The resulting graph is particularly useful for detecting the so-called "typical systems" of predictor values for classification or for nodes.

A classification tree, designed and built in accordance with our predictor's values, has also the ability to predict new tested lots of fuzes of hand grenades. Thanks to the developed tree's model and the introduction of new predictor values, the software automatically generates the value of the dependent variable, i.e., makes a post-diagnostic decision.

The software [7] has another tool that allows you to assess the quality of the built model, namely the predictive capabilities of this model. It is a graph of the cumulative value of the lift chart, which graphically summarizes the usability of the model to predict the value of the dependent variable in our case for the prediction of post-diagnostic decisions.

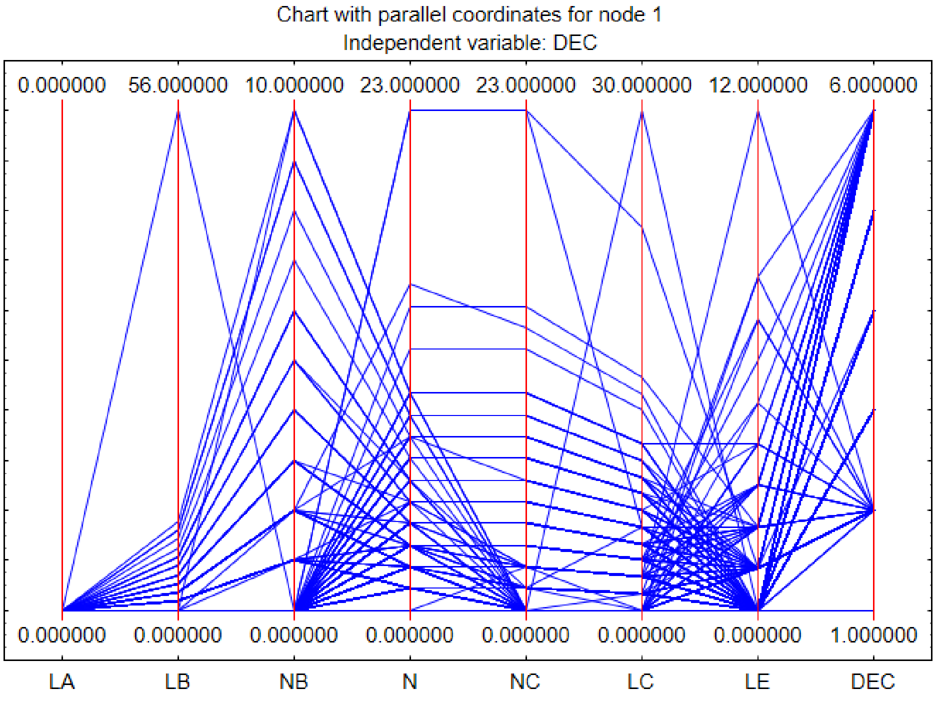

Fig. 7. The chart of data belonging to node 1

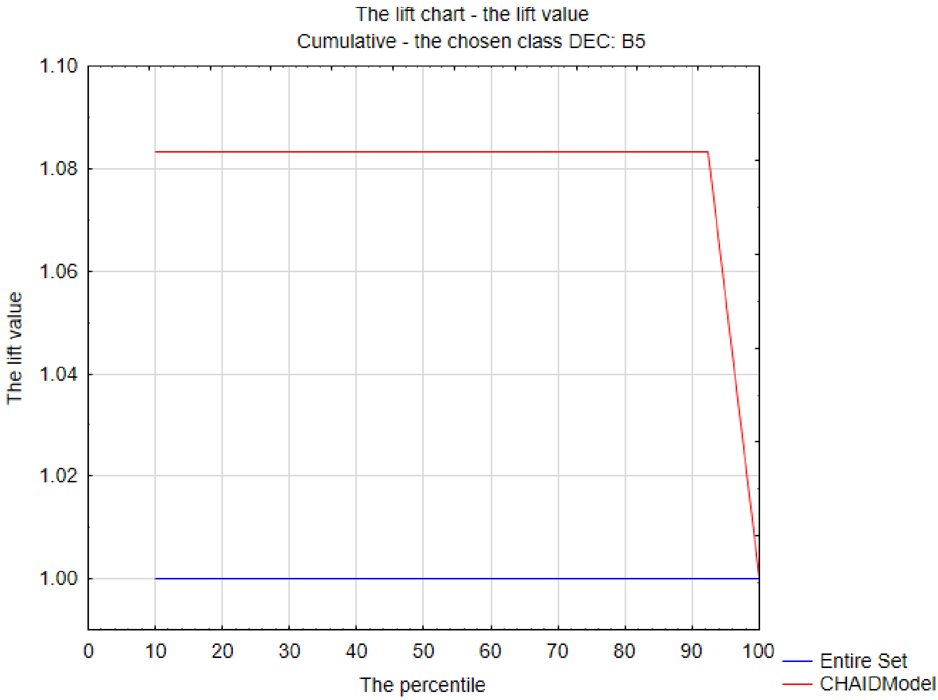

Fig. 8. The lift chart for "B5" decision

An example of such a graph of the growth value for the class "B5" is shown in Figure 8. These types of graphs were created for all possible values of the dependent variable. The axis (y) displays the values of growth, i.e., many times relative to the reference line, and on the axis (x) the values "percentile" are written, in other words, it is the value below which the values of a given percentage of the analysed samples fall.

This graph shows that in the range of cases from 10% to about 92%, most likely classified in the "B5" class, i.e., with the highest classification probabilities, we will receive a sample that contains about 1.08 times more cases than if the selection was of random nature. This graph is created for changing sets of observations that contain an increased number of cases with the highest probability of getting to a class, resulting from the model being built, so that the next one contains the previous one, and therefore this type of graph is called a cumulative one.

The tool used to assess the performance of the resulting model is the "ROC" curve (Receiver Operating Characteristic), which has many applications. It serves, inter alia, to assess the quality of the model predicting affiliation to various analysed classes. A useful indicator of the quality of the built model is the area under the curve.

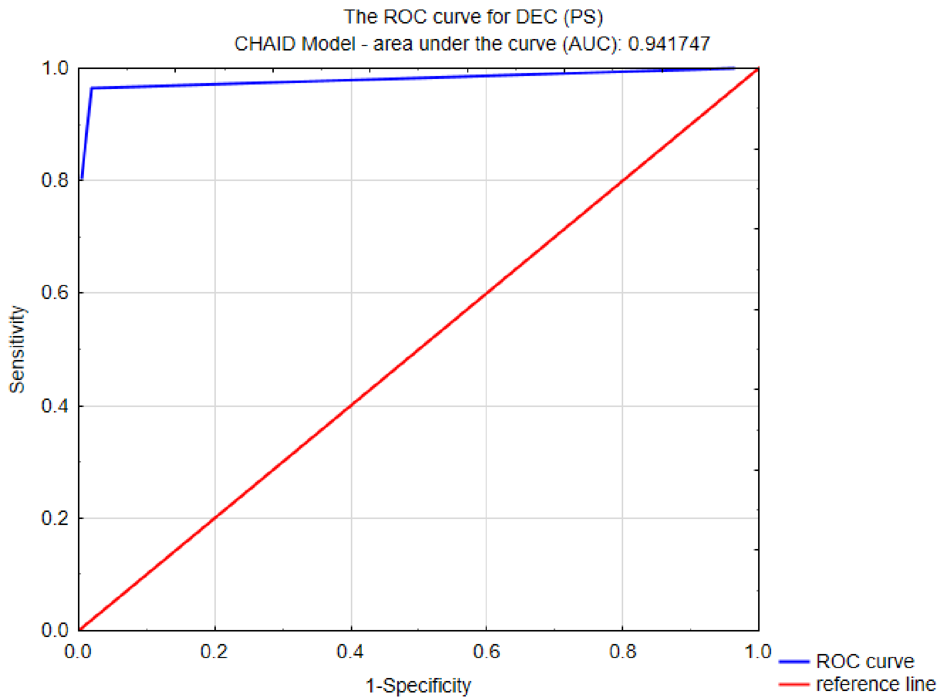

Fig. 9. The "ROC" curve for "PS" decision

The larger the area under the curve, the better the resulting model. An example of such a curve is shown in Figure 9, which shows that the area under the curve is 0.941747, which means that the built model in this class is very good. In order for the developed model to be a full model, "ROC" curves were created for all possible types of dependent variables analysed in our model, that is, for all post-diagnostic decisions.

5. CONCLUSIONS

The article attempts to describe the design and construction of a decision tree model according to the CHAID method for hand grenade fuzes of the UZRGM type. The prepared database of data results meets the detailed requirements of this method. The aim, set out at the beginning of the article, has been fully achieved.

During the design and construction of our decision tree, which in this case is a classification tree, due to the qualitative dependent variable, the necessary parameters were introduced which ultimately led to the development and selection of the best classification tree model. The resulting tree is relatively simple to build, but it contains in its formulas all possible analysed classes.

The specialised computer software, that was used to create our classification tree model, now allows us to automatically and quickly evaluate for new predictor values obtained from testing new lot of fuzes of hand grenades. Thanks to this designed model, the possibility of making human error while assessing data results is eliminated. The post-diagnostic decision, made by our model, for new data seems to be free of errors, which means that it was made in accordance with the applicable evaluation table of the tested UZRGM fuzes, which is included in the test methodology.

In recent years, we have seen a very intensive introduction of artificial intelligence to industrial processes and also to test procedures. Artificial intelligence in the form of, among others, decision trees will also be implemented in all types of assessment processes that take place when testing various technical objects. Such technical objects are, after all, among others hand grenade fuzes and other types of artillery fuzes. Tests of this type of technical objects will be carried out by research institutes and there will always be a need to assess their test results. Therefore, the need to correctly evaluate these test results becomes indispensable, and it is best to do it thanks to the developed evaluation model based, e.g., on decision trees and in our case on classification trees.

Of course, the implementation of this classification tree in practice requires, first and foremost, the approval of the management of the given research department and the connection of computers with new test results to our developed evaluation model in the form of a classification tree. Performing all these activities seems to be a very near future.

FUNDING

The article has been prepared in the frame of statutory activity of the Military Institute of Armament Technology in Zielonka.

REFERENCES

[1] Gatnar E. 1998. *Symbolic method classification of data* (in Polish). pp. 163÷164, 167÷171, 176÷178. Warsaw: Publishing House PWN.

[2] Łapczyński Mariusz, Tomasz Demski . 2019. *Data mining – prediction methods* (in Polish). Pp. 23-26. Statsoft Poland.

[3] Cards from laboratory tests of fuzes type UZRGM (in Polish). Archive of Military Institute of Armament Technology (MIAT).

[4] *Reports from tests of ammunition* (in Polish). Archive of MIAT.

[5] *The methodology of diagnostic tests of ammunition* (in Polish). Archive of MIAT. Index N-5001b.

[6] *Ammunition of land forces. Handbook* (in Polish). 1985. Ministry of National Defense, Chiefdom the Service of Armament and Electronics, pp. 224÷225.

[7] *Statistics 13.3 PL- computer software* (in Polish). 2018. Statsoft Poland.

Predykcja decyzji post-diagnostycznych dla badanych zapalników do granatów ręcznych w oparciu o drzewa decyzyjne

Dariusz AMPUŁA

Wojskowy Instytut Techniczny Uzbrojenia
ul. Prymasa S. Wyszyńskiego 7, 05-220 Zielonka

Streszczenie. We artykule przedstawiono krótką historię powstania drzew decyzyjnych oraz określono cel podjętych prac. Pokazano proces budowy drzewa klasyfikacyjnego według metody CHAID, zwracając szczególną uwagę na wady, zalety oraz cechy charakterystyczne tej metody a także na wymagania formalne, które są niezbędne do zbudowania tego modelu. Scharakteryzowano metodę budowy drzewa dla zapalników UZRGM, określając cechy badanych zapalników do granatów ręcznych oraz zastosowane predyktory, które są konieczne do tworzenia prawidłowego modelu drzewa. Zbudowano drzewo klasyfikacyjne na podstawie posiadanych wyników badań, przyjmując jako jakościową zmienną zależną przyjętą decyzję podiagnostyczną. Pokazano schemat zaprojektowanego drzewa dla pierwszych badań diagnostycznych, jego pełną strukturę oraz liczności poszczególnych klas węzła. Określono macierz błędnych klasyfikacji, która określa trafność błędnych predykcji, czyli poprawność dokonanej klasyfikacji. Przedstawiono arkusz z oceną ryzyka oraz błędem standardowym dla próby uczącej i v-krotnego sprawdzianu krzyżowego. Na wybranych przykładach oceniono jakość powstałego modelu predykcyjnego za pomocą wykresu skumulowanej wartości współczynnika przyrostu oraz krzywej "ROC".

Słowa kluczowe: inżynieria mechaniczna, drzewa klasyfikacyjne, gałąź, liść, węzeł, cecha