INNOVATIVE MECHANICAL ENGINEERING **ISSN 2812-9229 (Online)**
University of Niš, Faculty of Mechanical Engineering
VOL. 1, NO 1, 2022, PP. 147 - 157

Original scientific paper*

APPLICATION OF THE DECISION TREE IN VEHICLE PROCUREMENT DECISION MAKING

Bratislav Lukić[1,2], Goran Petrović[2], Predrag Milić[2], Dragan Marinković[3], Žarko Ćojbašić[2]

[1]Ministry of Defense of the Republic of Serbia

[2]University of Niš, Faculty of Mechanical Engineering, Serbia

[3]TU Berlin, Institute for Mechanics, Berlin, Germany

Abstract. *This paper discusses the problems of decision making when not all parameters for decision making are known, i.e. when future events must be predicted without the probabilities of events being known. Decision making under uncertainty has a very important role in today's business. Due to the lack of information and the unknown future state of the environment, decision makers find themselves in very ungrateful situations. The application of different decision making methods facilitates the decision making process. This paper describes the efficient application of the decision tree method to the vehicle procuring problem. A decision tree is a very good and simple tool which can expedite reaching an efficient solution and is extremely suitable for presentation due to its visual appeal. For the observed problem, a solution using the proposed decision tree method is presented under uncertainty during the procurement of vehicles. Calculation was performed to obtain a result, which is presented in graphical form.*

Key words: *Decision making, Uncertainty, Methods, Decision tree.*

1. INTRODUCTION

Making decisions is a part of everyday life and its very origin is connected to the origin of humanity. The first appearances of decision theory are closely related to the theory of organization and its founder, Frederick Taylor. The basis of decision theory is system analysis, which uses a systems approach as a scientific method [1]. Making decisions involves many of the processes that precede it, including gathering information, creating, observing, and evaluating alternative courses of action, as well as the implementation and evaluation processes that should follow the final decision.

*Received November 06, 2021 / Accepted February 03, 2022.
Corresponding author: Bratislav Lukić
Ministry of Defense of the Republic of Serbia, University of Niš, Faculty of Mechanical Engineering, Serbia
E-mail: lukic.bratislav@yahoo.com

Making decisions under uncertainty is done when several events are known that can happen after the decision is made, however, the probability of some of them happening is not certain. The decision made under uncertainty is done in cases when it is possible to determine future events, but not their probability. In order to make a decision i.e. make a choice among the alternative actions, it is necessary to predict the future. Making decisions presupposes predictions of possible outcomes or consequences of alternative actions, in various possible situations. Prediction of the state of affairs, i.e. the outcome of an action in the future, is based on experience and knowledge, that is, information the decision maker acquired in the past. There is no way to predict future events in all segments, except for some statistical methods where the future can be inferred to a greater extent based on the past knowledge. However, uncertainty is also present when a decision is made with all decision parameters known. Certain criteria of the value of the qualitative nature of attributes for the evaluation of alternatives depend on the subjective assessment of decision makers, as well as the relative weight of selected criteria. Although the subjectivity of decision makers cannot be avoided when solving real problems, uncertainty must be accounted for during the decision making process. Better knowledge of the decision making process under uncertainty affects the improvement of choices and minimizes the possibility of making the wrong decision. The problem of decision making is that it is necessary to choose one of the available alternative actions, not knowing which of the uncertain events will occur, and the consequence of the action taken depends on which uncertain event will occur. Individual events have more possible actions and thus more possible consequences. There are methods that allow the decision maker to choose the optimal action and they differ greatly. The decision maker can apply and use them at all stages of the decision making process, depending on one's own expertise and desire to use them. Quantitative methods are mainly used in comparing and evaluating alternatives. However, numerical data are useful if the previous assumptions are true. Quantitative methods in most cases not only participate in comparing and evaluating alternatives, but also define them. In the decision making process, a large number of variables cannot be quantified and there are also a large number of factors that would have a significant impact on the consequences of decisions, which fall outside the reach of the decision maker or are unavailable within given time and cost limits. In decision making, probability serves as a means of expressing uncertainty. In most cases, based on their experience as well as the relative frequency of events, the decision maker draws conclusions from partial information grounded in one's own intuition. Probability theory provides useful information, because it reduces the degree of uncertainty for the decision maker, helps in comparing and evaluating alternatives, as well as in preparing for the choice. The values of the probability of occurrence of individual events are mostly subjective, although they are based on objective, accessible evidence, because they include the experience of the decision maker, as well as one's readiness, value, personality, observations and preferences.

Most of the decisions made contain a certain risk. In addition to certain levels of risk, the decision maker must have a certain analytical framework, within which one can reduce the uncertainty after the decision is made to a level that allows the consequence that satisfies the decision maker in the choice of goal to come to the fore. The analytical framework can be a utility matrix, a decision tree (DT), or more recently a fuzzy set. When applied appropriately, these methods can assist the decision maker in analyzing alternatives as well as in evaluating possible consequences. The technical procedure for making decisions under risk and uncertainty is a DT. It is used when a decision maker needs to consider a whole series of decisions at the same time. The method was named the decision tree, because of a tree-structured diagram that can show all possible solutions to a problem and all possible outcomes of an event or classify certain data. Each branch of the DT can usually be represented as a single "*if-then*" rule.

The DT as a machine learning method is not new and is successfully applied in many areas. This paper presents the possibility of applying DT in the field of procurement logistics. The application of a DT is shown on a real problem with real data and restrictions defined by the decision maker.

2. State-of-the-art

DT is one of the machine learning techniques that presents all possible alternatives and follows each alternative to its conclusion, where an easy comparison can be made between different alternatives [2].

Graphical representation is one of the best advantages of the DT, because it is easy to understand compared to other classification algorithms. This technique is used for both continuous and discrete data sets. The most used DT algorithms are ID3, C4.5, and CART [3]. However, they are only used for small data sets and all or part of the entire data set needs to remain permanently stored. This problem is solved by the DT algorithms SPRINT and SLIQ [4].

Data uncertainty is common and can be caused by various factors, such as limited measurement accuracy, outdated sources, etc. In [5], an improved traditional DT algorithm is presented, where entropy and information retrieval are included, accounting for the uncertain data interval and the probability distribution function.

The study [6] considers the problem of classification with imperfect data. That is, standard DTs are expanded with uncertainties in construction and classification procedures. Uncertainty is represented by the possibility of distribution. The extended approach allows for a more flexible construction process by allowing the selection of more than one attribute in each node.

The DT has found application in multi-decision tables as a means of presenting knowledge. Where the goal of the study [7] is to build trees with a reasonable number of nodes and optimal accuracy, because it is not enough just to consider the global rate of misclassification of the DT, it is necessary to study the local rate of misclassification. It may be that the global rate of misclassification relative to the entire DT is small, however, the local rate of misclassification related to the terminal nodes of the tree may be too high. The novelty of this paper is in the application of DT in decision making under conditions of uncertainty, where the aim is to reduce classification errors, both local and global.

3. Decision tree

Graphical representation plays a major role in modeling in a structured decision problem. The most used impact diagram is an acyclic, directional graph in which each node indicates one variable of the decision problem, while arcs represent the relationships between the variables. Another graphical representation is the DT, which significantly facilitates decision-making, because the consequences of the decision are forked. Using this technique shows the development of future events as they would happen in the future. In the case of a DT, decisions and uncertainties are first listed chronologically and then the tree is constructed in order to show the possible outcomes of alternative decisions. Each branch exiting the decision node corresponds to one alternative, each branch exiting the random node corresponds to a possible outcome [8].

The formation of the DT practically leads to the formation of a chain of interconnected and dependent decisions that participate in the creation and making of the final decision. The DT is particularly suitable for situations where decision making consists of a series of smaller decisions that are timed one after the other. When using this decision making method, there are several parts that are known, such as the possibilities among which a decision must be made and the consequences that could occur if an option is chosen. It can be said that the mentioned method starts with a decision point, followed by alternative decisions that can be made and each of them is presented with its own branch that emerges from the decision point. Practically, the DT represents a map of possible solutions in different phases that are available to the decision maker. When making decisions the branch leading to the node of the case with the highest expected value is chosen [9]. What sets this method of decision making apart from others is the possibility of returning to the initial phase when it is noticed that the previous decision was made incorrectly and does not yield the desired result. The DT is a graphical interpretation of decision making analysis without sampling and decision making analysis with sampling. The above method of decision making in the case of single sampling is a transitional stage towards sequential decision making, as additional information is collected through one sampling and Bayes' theorem is applied. Sequential decision making involves performing a series of consecutive samplings until the decision maker is at some point satisfied with the information collected or until the cost of sampling outweighs the benefit of sampling. This necessitates a stop rule, i.e. sampling should not be performed if there is an action on any part of the tree whose expected regret is less than the sampling costs. This way, branching on that part of the tree is stopped and other parts are checked, until this criterion is met for all branches [10].

3.1 Decision tree algorithm

Fig. 1 presents an example of a DT, which shows the structure made of decision nodes marked with squares, possible consequence nodes marked with circles, and final nodes marked with a triangle. Decision trees are considered one of the most popular approaches to classifying data. The DT starts with a root, i.e. a node that has no input branches, while all other nodes have exactly one input branch. An output branch node or internal node divides the input data into two or more subspaces according to certain discrete functions of the input values of the attribute. In the simplest and most common case, each data accounts for one attribute, so that the data space is divided according to the value of the considered attribute. Each tree leaf is associated with a class and represents the best appropriate target value. Assigning a class to some data is done by going through the tree starting from the root node to the leaves. When creating, the selection of attributes for each node is done by heuristic methods. They are based on the assessment of the quality of discrimination of a subset of examples from the defined set, remaining for discrimination in the observed node. However, although a tree can excellently classify all examples from a defined set, this does not guarantee high accuracy on new examples, as they are often over-tuned to defined examples, which leads to simplifying, resulting in smaller trees that are both more accurate and more understandable [12]. The complexity of the tree has a key influence on its accuracy. It is explicitly controlled by the stop criterion used and by the applied pruning method.

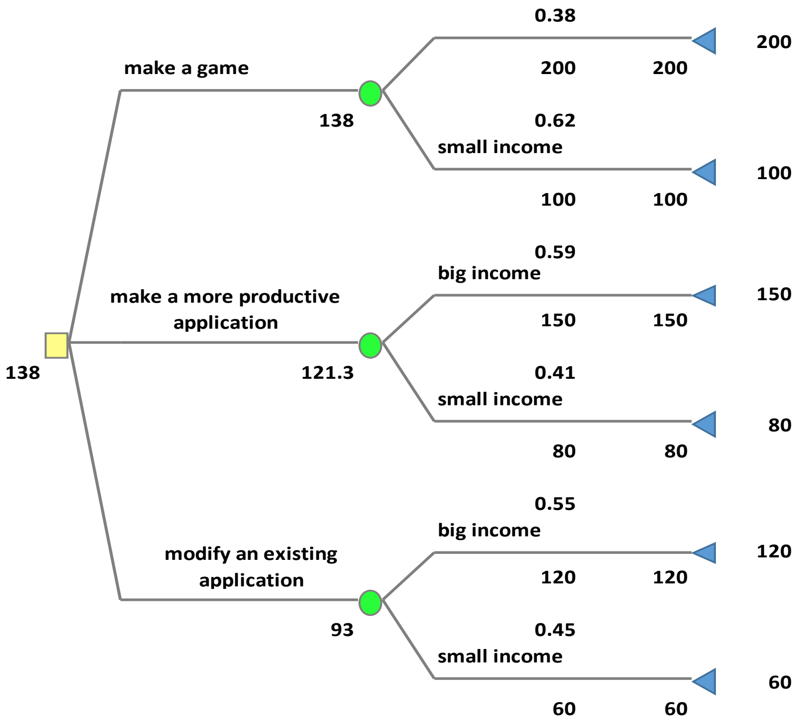

Fig. 1 Example of a DT [11]

Mechanisms for forming a DT are algorithms that automatically build a DT based on a given set of data. In most cases, the goal is to find a DT such that the classification error is minimal. However, other target functions can also be defined, such as the minimum number of nodes that a tree must have or the minimum average depth that a tree should reach. Representing an optimal DT based on the data is considered a difficult task, which is why heuristic methods are necessary to solve the problem of building an optimal DT. Two conceptual phases are significant: growth and pruning. In each iteration, the algorithm considers partitioning a data set to form a tree using the result of a partitioning function on a discrete data set, which is applied to the input attributes. The choice of the most suitable function was obtained according to some division measures. After selecting the appropriate division, each node further divides the tree-forming data set into smaller subsets, until the measure value for each of the divisions is less than the specified measure value that must be met for the division to occur or for the stop criterion to be met. In most cases, the partition functions on a discrete data set dividing the internal node according to the value of one attribute. Consequently, the mechanism for building the DT finds the best attribute by which the division will be made. For the criterion of dividing by the values of several attributes, several attributes can participate in one division node.

Searching for the best division criterion by the values of multiple attributes is more complicated than searching for the best division by the value of one attribute. Although this type of criterion can significantly improve the tree performance, these criteria are much less popular than single-attribute value splitting criteria. Most criteria for dividing by the values of multiple attributes are based on a linear combination of input attributes.

The growth phase of the tree continues until the stop criterion is met. Applying a strict stop criterion tends to create small DTs and DTs that are too generalizing. On the other hand, the use of a loose stop criterion leads to the creation of a large DT, overcrowded in relation to the set of initial decision data. To solve this dilemma - "loose" or "strict" criterion, undercutting methods have been developed [12]. A crowded tree is transformed into a smaller tree by removing sub-branches (sub-trees) that do not contribute to the reduction of accuracy. It has been shown in various studies that the application of pruning methods can improve the performance of the DT. Another major motive for undercutting is "substituting accuracy for simplification" [13]. When the goal is to make a sufficiently precise compact description of the terms, undercutting is very useful. The accuracy of the trimmed DT indicates how close it is to the original tree. There are several techniques for pruning the DT. Many of them are applied using movement through the nodes from top to bottom or from bottom to top. The most popular undercutting techniques are undercutting the complexity of the cost, undercutting the reduced error, undercutting the smallest error, pessimistic undercutting and error-based undercutting. Several studies have tried to compare the performance of different trimming techniques and the results show that some methods (such as the trimming cost complexity and the trimming reduced error) tend to overcut the tree, making a smaller, less accurate DT. Other methods (such as the error-based pruning method, the pessimistic error undercutting and the least error undercutting) tend to undercut the tree insufficiently. However, the general conclusion is that there is no method of undercutting that dominates over the others.

3.2 Software tools for creating a decision tree

Simpler DTs can be drawn manually on a sheet of paper. However, in order to save time and reduce the risk of manually creating a DT, many tools have been developed to create a DT. They are especially useful when it comes to complicated problems that need to be decided. Simple tools for word processing and spreadsheets such as MS Excel can serve this purpose, but there are many specialized tools for creating a DT. The process of creating a tree is very fast and efficient when the user becomes proficient in using the selected tool. However, one should know that the tool is as good as the user of the tool, because the tool does not make a decision, but only follows the user's instructions and it is up to the user to learn how to give good instructions. Here are some of the tools for making a DT:

- **Insight Tree** is easy to use if the user is already familiar with the concept of the DT, as it allows to add branches quickly, optimize nodes, and use different styles [14].
- **TreePlan** is one of the most famous tools for applying the DT method. It is an add-on for MS Excel and is used to draw and calculate the value of alternatives. It was developed by Michael R. Middleton and modified by James E. Smith. The tool has many advantages, such as: saving time when creating a DT model, very simple installation and implementation of analysis using tables in MS Excel [15].

- **Lumenaut** is used as a built-in add-on to MS Excel. In addition to the construction of a DT, it provides other tools for statistical analysis such as Monte Carlo simulation [16].
- **DT Analysis** (Vanguard Studio) is a platform add-on developed by Vanguard Software. It allows the creation and analysis of a DT. After creating a DT model, it offers some additional options that can be performed, such as defining risk appetite when analyzing decision models; creating tables and graphs that allows viewing of all the possible outcomes and probabilities; calculating the prediction value [17].
- **SmartDraw** is a commercial tool that allows quick and easy drawing of different types of diagrams, including DTs. The tool contains a large database of examples for each type of diagram [18].

4. Application of the DT approach

Based on the previous section, it can be concluded that DT is a very applicable method of decision making. Some areas of application are economics, logistics, medicine, agriculture, construction, etc. In these areas, the method is used mainly for some more complex analysis. Considering that the selection of the optimal means of transport during procurement considers attributes that are extremely complex and influenced by various overlapping factors, DT was chosen as it is a well-known and often used machine learning technique for finding internal relationships in data. Moreover, the significant influence for the choice of the DT method is exerted by the uncertain conditions in which the decision is made, having in mind the unknown probability when choosing the offered alternative.

The main means of passenger transport is the bus, which is quite expensive, given that the number of passengers who will need the services in question is uncertain. In order to make decisions about expanding or improving transport services the DT method can be extremely useful. The paper will use data obtained for Mercedes-Benz and Scania vehicles. The aspiration of every company is to achieve an increase in profits. In such a case the decision will be made whether to purchase a new bus, a used bus or keep the existing condition. A study on 168 buses by classes with the same or similar characteristics was considered. When solving the problem, the values of buses given in [19-22] were used, where buses with basic and additional equipment were considered, i.e. in addition to the stated characteristics, the condition from the aspect of wear was considered for used buses. The price of the service of the new buses was taken over from the authorized services for Mercedes-Benz and Scania, while the average maintenance costs of an approximately 10-year-old fleet were used for the maintenance of used buses, as the buses of that age were considered. When classifying used vehicles according to the state of wear, the main factor was the number of kilometers traveled.

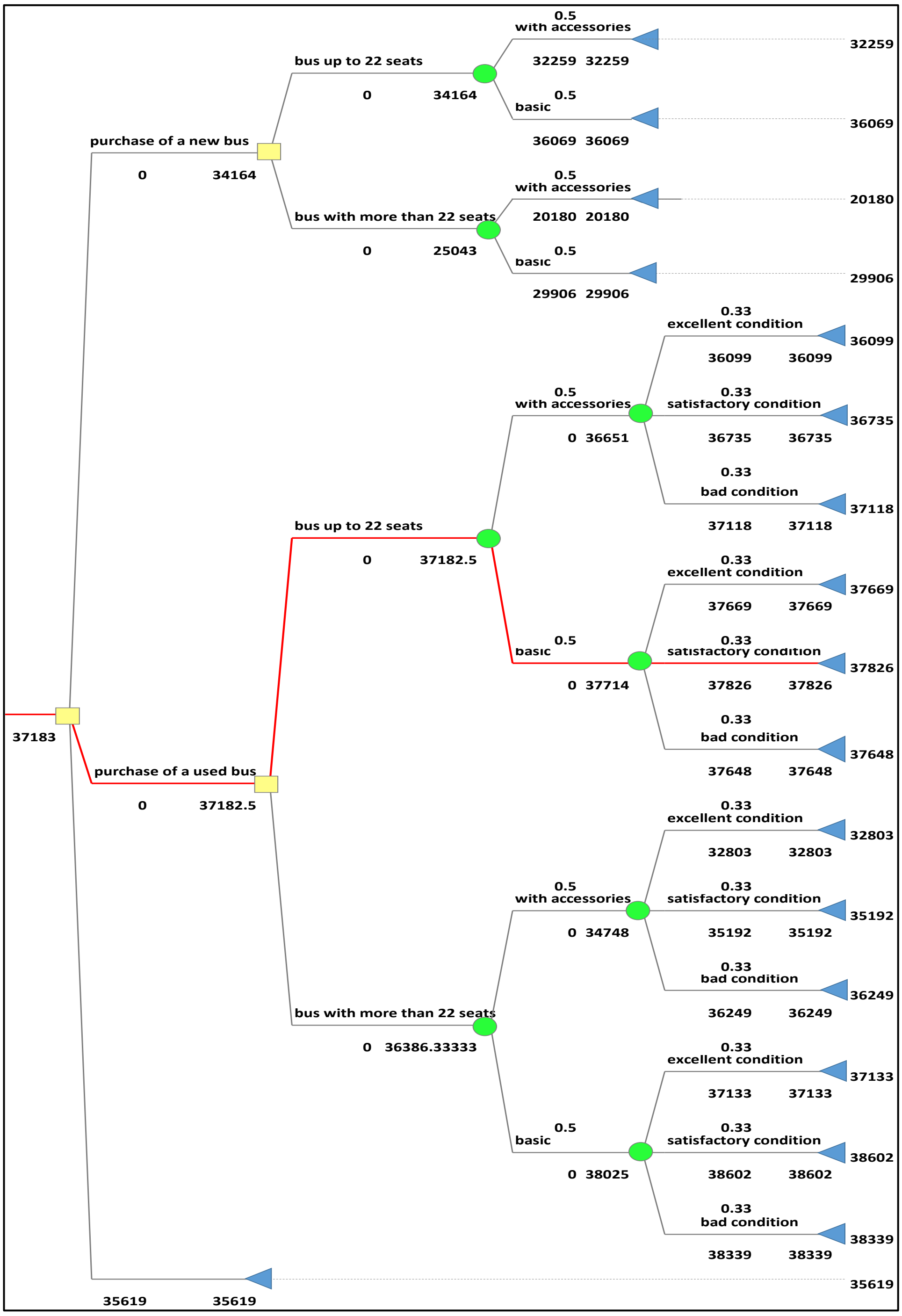

Fig. 2 DT when choosing the procurement of a bus - minibus

The classification was performed according to the following:

- buses with up to 22 seats - excellent condition (average number of km traveled around 150,000), satisfactory condition (average number of km traveled around 300,000), poor condition (average number of km traveled over 500,000);
- buses over 22 seats - excellent condition (average number of km traveled around 300,000), satisfactory condition (average number of km traveled around 700,000), poor condition (average number of km traveled over 1,000,000);

Since the problem is considered in terms of uncertainty, the probability of the number of passengers who need a transport service was not considered, but the bus capacities were calculated. For buses with 22 seats, 22 passengers were considered, and 46 passengers were considered for buses over 22 seats. Since the main expense when using vehicles is fuel, the fuel consumption for the distance of 250 km and the number of motor days at work in one calendar year were considered. This research estimates the gained profit in one calendar year in the amount of 35,619,000 RSD.

In Fig. 2, one can see that the best choice is the second alternative, the purchase of a used vehicle, i.e. a bus with up to 22 seats with basic equipment and satisfactory condition, because the purchase of this bus makes the largest profit. The branch that corresponds to the best alternative is shown in red from the initial node - the root to the final node - the leaf.

The result was obtained based on the predefined input data limits, where the probability of consequences was not considered, because the decision is made in conditions of uncertainty. However, stability can be found in this study, which is still essential for vehicle procurement. However, as the difference in the values of the possible consequences is small, there is a significant probability of errors in determining the potential input values. This implies that it is necessary to perform a good analysis when defining the constraints for the selection of input data.

5. Conclusion

Decision making is a concept that is especially important in the business world. Managers are daily confronted with the concept of decision making, i.e. making various decisions. These decisions can range from the simplest to the most complicated. When it comes to complicated and complex decisions, they are mostly decisions that are made under uncertainty, but also under risk. Apart from some statistical methods where the future can be inferred based on knowledge of the past, there is no way to predict future events in all segments. The decision maker assumes certain conditions based on the collected data, which in most cases are accurate and slightly variable as a function of time. Making decisions under uncertainty is a very common term, but it can be very dangerous for business, especially because it can result in various losses and mistakes. However, uncertainty is also present when making decisions with all known parameters for decision making.

The decision tree is a very applicable method for making decisions when choosing alternatives in the procurement of goods or services, due to the applicability in conditions of uncertainty and risk. This method graphically shows the possible decisions and their consequences and is an extremely important application in situations where it is necessary to make several decisions in a row. The method starts with the initial node, i.e. the problem or decision that needs to be solved and then the nodes of decisions and nodes of consequences are defined, i.e. the branches of alternative activities and the branches of possible consequences. The DT consists of three steps. In the first step a logical model of the tree with all the data is built, followed by backward algorithm calculation and finally the optimal way is found by forward counting. However, like all other methods, the DT has certain advantages and disadvantages. The main advantages of this method are transparency, efficiency, graphical presentation and training speed, while the most significant disadvantages are complexity, required education, experience and small dataset.

The presented application of DT in procurement logistics enables the most objective decision to be made based on the real data and conditions. However, it is necessary to emphasize the need to correctly determine the criteria for selecting offered alternatives, because certain qualitative criteria of the attributes value for the evaluation of alternatives depend on the subjective assessment of decision makers, as well as the relative weight of the selected criteria. The subjectivity of decision makers cannot be avoided when solving real problems, but uncertainties should be accounted for during the decision making process. In the presented example, it was concluded that a used bus with up to 22 seats with basic equipment and satisfactory condition is the best solution.

Acknowledgement: *This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-9/2021-14/200109) and by the Science Fund of the Republic of Serbia (Serbian Science and Diaspora Collaboration Program, Grant No. 6497585).*

REFERENCES

1. Hansson, S.O., 2005, *Decision Theory: A Brief Introduction. Department of Philosophy and the History of Technology*, Royal Institute of Technology (KTH), Stockholm, pp. 26-28.
2. Jijo, B.T., Abdulazeez, A.M., 2021, *Classification Based on DT Algorithm for Machine Learning,* Journal of Applied Science and Technology Trends, pp. 20-28.
3. Patel, H.H., Prajapati P., 2018, *Study and Analysis of DT Based Classification Algorithms,* International Journal of Computer Sciences and Engineering, pp. 74-78.
4. Priyam, A., Abhijeeta, G.R., Guptaa, R., Ratheeb, A., Srivastavab, S., 2013, *Comparative analysis of serial decision tree classification algorithms*, International Journal of Current Engineering and Technology, pp. 334-337.
5. Qin, B., Xia, Y., Li, F., 2009, DTU: *A decision tree for uncertain data*, Springer, Advances in Knowledge Discovery and Data Mining pp 4-15
6. Jenhani, I., Amor, N.B., Elouedi Z., 2008, *Decision trees as possibilistic classifiers*, Elsevier, International Journal of Approximate Reasoning, pp.784–807.
7. Azada, M., Chikalovb, I., Moshkovc, M., 2020, *Representation of Knowledge by Decision tree s for Decision Tables with Multiple Decisions*, Elsevier, Procedia, pp. 653–659.
8. Gilboa, I., 2011, *Making better decisions: decision theory in practice*, John Wiley & Sons, p. cm.
9. Sikavica, P., Hunjak, T., Begičević Ređep, N., Hernaus, T., 2014, *Business decission making*, Textbooks of the University of Zagreb, School book, Zagreb
10. Čupić M., Tumala M., Suknović M., 2008, *Decision making - formal approach*, Faculty of Organizational Sciences, Belgrade
11. Lucidchart. *What is a decision tree diagram* [Online], Available at: www.lucidchart.com/pages/decision-tree (last accessed: 18 Apr 2021)
12. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984. *Classification and Regression Trees*, Wadsworth, Belmont, p. 368.
13. Bratko, I., Bohanec, M., 1994, *Trading accuracy for simplicity in DTs*, Kluwer Academic Publishers, Boston, pp. 223-250.

14. Insight Tree. [Online]. Available at: www.visionarytools.com/index.htm (last accessed: 22 Apr 2021)
15. Treeplan® DT Add-In for Excel. [Online]. Available at: www.treeplan.com (last accessed: 22 Apr 2021)
16. Lumenaut a built-in add-in to *MS Excel*. [Online]. Available at: www.lumenaut.com (last accessed: 22 Apr 2021)
17. Supply chain planning. [Online]. Available at: www.vanguardsw.com/decision-tree-analysis-add-on (last accessed: 22 Apr 2021)
18. DT Project development. [Online]. Available at: www.smartdraw.com (last accessed: 22 Apr 2021)
19. Official site in Serbia for Mercedes - Benz - The standard of buses. [Online]. Available at: www.mercedes-benz-bus.com/sr_RS/home.html (last accessed: 29 Apr 2021)
20. Official site in Serbia for Scania. [Online]. Available at: www.scania.com/rs/sr/home/products-and-services.html (last accessed: 29 Apr 2021)
21. The first Serbian site dedicated to buses. [Online]. Available at: www.autobusi.net (last accessed: 29 Apr 2021)
22. Online platform for selling vehicles. [Online]. Available at: www.autoline24.rs (last accessed: 29 Apr 2021)