substantiated and chosen. A testing system has been developed, which allows modeling, debugging, modernization of the mathematical apparatus, provides full access to each element and the possibility of complete change and modernization. UML diagrams of usage options, classes, components, and deployments were built at the system design stage. The software is developed using neural convolutional networks and using the Python programming language and modules Matplotlib, numpy, Keras, PIL, OS. The neural network was trained on a set of test photos. Proven effectiveness of the proposed technology on a real object - test images of car logos. A number of experiments were performed using different optimizers and the number of filters, which allowed to identify the optimal settings of the neural network. It was found that the best result is achieved when using the SGD optimizer, with the number of epochs should be approximately 200. It has been experimentally established that certain settings allow you to perform recognition of car brands with an accuracy of 80-95%.

Keywords. *Neural network, artificial intelligence, convolutional neural network, algorithm, identification, recognition, photographic image, Python, Keras, software, cars.*

Стаття надійшла 22.12.2021 р.

УДК 004.02 **doi.org/10.31498/2522-9990252023286733**

Kotykhova L. D., Levitskaya T. A.

APPLICATION OF DECISION TREE METHOD FOR CLASSIFICATION PROBLEMS IN DECISION SUPPORT SYSTEMS

In any field of human activity can be traced the significant impact of information technology. Although people create very complex constructions in various fields that no computer can do, they lose when they have to solve repeating tasks with the same result. In such tasks the human factor and the presence of errors are inevitable, which can be avoided by automating the process of solving such problems and entrusting this work to a computer.

The rapid development of information technology has led to the emergence of decision support systems (DSS), which greatly facilitate the work of people in many different areas of activity, where there is a need for decision-making by employees.

Of particular importance is the use of decision-making systems to process people's applications in various institutions. Such problems are mostly related to classification problems. There are several methods for solving them, but one of the most well-known is the decision tree method.

The article is devoted to the analysis of the application of the decision tree method for classification problems indecision support systems. The article describes the object of research, systematizes the known approaches to solving the problem, and discusses the advantages of using the decision tree method in decision support systems in general and, in particular, for processing people's applications in various institutions.

The use of the decision tree method for solving the problems of classifying applications will help to significantly improve the quality of customer service of institutions, make the process of processing applications and waiting for a decision on them more comfortable, and, accordingly, will improve the quality of life of the population.

Keywords. *Data mining, decision tree method, decision support system, classification problems.*

Formulation of the problem. Along with the rapid development of information technologies, the pace of their introduction into institutions of various spheres of human activity remains low. Many tasks that could be performed by a computer are still performed by humans. Because of this, the speed of customer service in institutions remains low. Therefore, visiting various institutions, especially where there is a need to submit any applications, is an additional stress for people and is associated with a number of difficulties. For example, applying for a subsidy or for a loan takes people a lot of precious time and effort.

In a situation where workers have to perform a significant number of identical tasks in the shortest possible time, the human factor can play its role. Under such working conditions, people get tired quickly, their attention and concentration, and therefore work results, deteriorate significantly. Such features can lead to wrong decisions and various errors. It is also critical that with such work, the result of the same tasks can be different every time, which is unacceptable when making important decisions.

The use of decision support systems to automate the processing of such applications is an urgent task. The introduction of decision support systems for processing customer applications will significantly reduce the processing time for applications and the time spent in institutions, will help reduce queues in institutions, which is especially important during the period of quarantine restrictions. And all this, in turn, will significantly improve the working conditions of employees of institutions and lead to an improvement in the emotional atmosphere.

Analysis of recent research and publications. Today, active scientific and practical research is being conducted on the formalization and computerization of decision-making tasks in various fields of human activity. General approaches to solving such problems are described in particular in the works of the following authors: Y. Petrunya, P. Bidyuk, L. Korshevnyuk, *J.* R. *Quinlan,* J. Rasku, P. Kuusipalo, H. Joutsijoki, L. G. Kabari, E. O. Nwachukwu, Yan-yan Song, Ying Lu, Pang-Ning Tan**,** M. Steinbach**,** A. Karpatne**,** V. Kumar**, T. Mitchell,** S. Neelamegam, E. Ramaraj.

In their work J. Rasku, P. Kuusipalo, H. Joutsijoki say, that a common feature of all decision support systems is the fact that they combine information from several aspects of the problem under consideration. And they believe that although the system logic may be simple, a person cannot remember all aspects at the same time to make a decision [1].

In the work [2], the author considered the classification of information signals based on the optimal structure of the binary wavelet transformation tree. This method ensures the necessary accuracy and reliability of classification, even if there are noise components in the information signals.

The authors of the work [3] call ensembles of decision trees one of the most effective classification tools. In their work, when conducting experiments to compare algorithms for building ensembles of decision trees, they established that the best option can be achieved using **the Extra-Trees algorithm** for building an ensemble of trees. In classification, this algorithm provided the highest accuracy on seven data sets out of ten.

The aim of the study. The main goal of this work is to study the possibility of using decision support systems to process various customer applications in various institutions (for example, banks, social service centers) and to study the application of the decision tree method to solve classification problems in decision support systems.

The main research material. Deciding on customer applications is time consuming, which leads to inconveniences and problems for people, such as applying for subsidies or loans.

The object of this research is the decision-making process when considering appeals from clients of various institutions, using the example of a bank and the issue of issuing loans and the center of social protection of the population with the issue of assigning benefits or additional payments.

Such tasks involve processing a huge amount of data and documents about each client. Often, such processes are not automated or not automated enough, which is why they employ a large number of workers. Therefore, the main problem of such tasks is the high labor intensity of this work. It is not uncommon for several departments and employees to assess, for example, a client's creditworthiness at the same time.

Today, very often employees of social protection departments fail to meet the statutory deadlines for calculating subsidies. As a result, people are forced to pay the full cost and then recalculate, which is an additional burden on both subsidy recipients and employees of social security departments. This situation is created due to the fact that the employees of the departments do not have time to process a large amount of information manually [4].

A similar situation is observed in banks in the process of processing loan applications. People who need money have a long time to wait for their applications to be processed which leads to additional problems for them.

In general, situations in which it is necessary to make decisions occur with great frequency in various spheres of life, in various systems, processes and phenomena. Deciding whether or not to accrue subsidies and loans can be significantly accelerated by automating this process. DSS is used to help and facilitate the process of making informed decisions and reducing the impact of the human factor.

In general, a decision support system is a computer-based, automated information system that helps people make decisions in difficult situations [5].

When using decision support systems for processing customer applications, it is most often necessary to solve the classification problem.

The classification is called the method of data mining (machine learning) used to predict group membership for data instances. The classification process involves breaking down many objects into classes according to certain criteria [6].

Thus, the task of processing applications in a bank or center of social protection is reduced to the dividing of all applications by groups: to confirm the issuance of money or to refuse. To perform classification, the set of initial data (data sample) must be divided into two parts: the training set, which will be used to construct the model, and the test set, which is used for testing the model. These sets contain input and output (target) values. The output values of the training set are intended for training the model, the test set - for checking the performance of the model.

Different methods are used to solve the classification problem. S. Neelamegam in his work names «the main ones:

- Decision Tree,
- K-Nearest Neighbor,
- Support Vector Machines,
- Naive Bayesian Classification,
- Neural Networks» [6].

To solve the scientific problem, the use of decision trees was chosen. The decision tree method is used in a wide range of human activities, especially in automated decision making. Classification using decision trees is one of the most common methods used to solve this problem. For the classification task, the decision tree method consists in the process of dividing the raw data into groups until homogeneous or nearly homogeneous sets are obtained.

The main area where decision trees are used is the processes of making management decisions, used in statistics, data analysis and machine learning, which is widely used in the creation of artificial intelligence, as are neural networks. Decision trees are also used in operations research, particularly in decision analysis, to help determine the strategy most likely to achieve a goal. It is commonly used

to solve a classification problem because the tree is easy for the user to understand and make decisions with a schematic representation of the problem.

When processing applications for a subsidies or loan, the criteria that need to be taken into account for making decisions are known in advance. For example, does the client have transport, real estate, does the area of his house exceed a certain value, etc.

A number of questions and possible answers to them can be organized in the form of a decision tree, which is a tree-like hierarchy consisting of nodes and directed edges [7]. Moreover, since each question can be answered “Yes” or “No”, branching from each node will only go in two directions.

Tree-like structures consist of rules of the form “If ..., then ...”, which are created in the process of learning using the training set.

When building and training a decision tree, which will be used to solve a specific problem, it is necessary to work out the following steps in sequence:

- First of all, it is necessary to choose an attribute that will be used to divide the set into subsets in a specific node. This attribute is also called the partition attribute.
- After that, at the next stage, you should choose the criterion that will be used to stop the study.
- If necessary, choose methods that will simplify the decision tree, that is, cut off certain of its branches.
- When the tree is built, it is necessary to provide an assessment of the accuracy and correctness of its work.
- In the nodes of the tree there are rules, and here it is checked whether this rule corresponds to the attributes from the training set. As a result of such a check for a binary tree, all examples that have entered a node are divided into two subsets, one of which contains examples that satisfy the rule, and the other contains examples that do not. If the tree is not binary, it is possible to divide the examples into more than two subsets. One independent variable is tested at the nodes of decision trees. If it can accept categorical values, then for each possible value a new branch will emerge from the node of the tree. If the variable is a numeric value, then it must be compared with a constant value. Values of dependent variables (classes) are recorded in the leaves of the tree. After dividing the initial set into subsets as a result of checking the first rule, the checking rule is again applied to each subset. This procedure is repeated recursively until some stopping condition of the algorithm is reached.

A tree can have three types of nodes:

- A root node that has no incoming edges.
- Internal nodes, each with one incoming edge and two outgoing edges.
- Leaf, each with one incoming edge and no outgoing edges [7].

That is, for the training sample A and a set of classes B, the following three situations are possible:

1. The set A consists of objects belonging to different classes. In this case, it is necessary to further divide the set A into subsets.
2. All the objects of the set A contain belong to the same class B. Then the decision tree for A is a leaf that defines this class;
3. The set A may not contain objects, that is, it may be empty. Then it is a leaf again, and it is necessary to move on to processing another set.

After all checks have been performed on the last node, the check and split is not performed because this node is a leaf. The leaf determines the solution for each example that falls into it. Thus, unlike a node, a leaf does not contain a rule, but a subset of objects that meet all the rules of the branch that ends with this leaf. Obviously, to get into a leaf, an example must meet all the rules that lie on the way to that particular leaf. Since the path in the tree to each leaf is unique, each example can fall into only one leaf, which ensures the unity of the solution.

Figure 1 shows a fragment of the decision tree for the task of classifying customer applications.

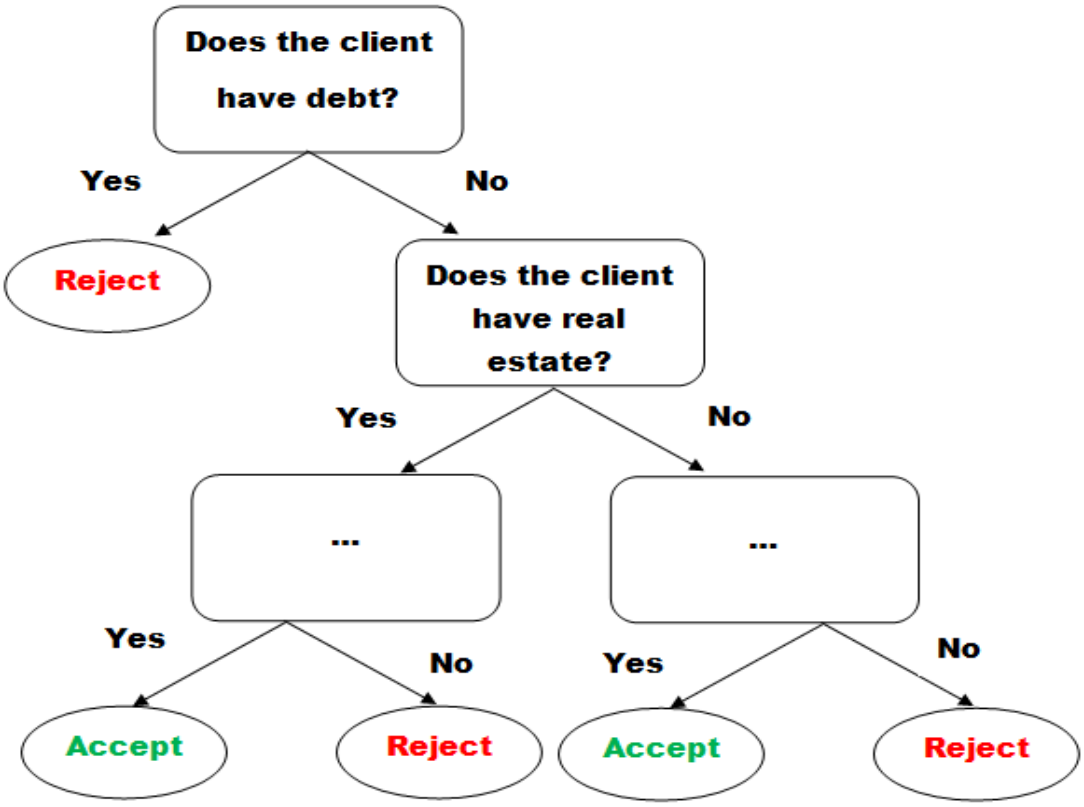

Figure 1 – A fragment of the decision tree for the task of classifying

As can be seen from this figure, the root and internal nodes of the tree contain different questions. Two edges come out of them: a positive and a negative answer to the question. The leaf contains the decision that must be applied to this application (accept or reject). Thus, passing through such a tree from the root to the leaf for processing the application, the necessary criteria will be taken into account and the decision will be automatically obtained. That is, the tree helps to classify all applications into two classes. The decision-making process itself is clear, and if necessary, it can be traced, due to non-compliance with which requirements, a particular client was denied.

The quality of the constructed tree depends on the choice of variable used for initial partitioning. This variable should divide the set so that the subsets contain objects of the same class, or objects as close as possible to it.

To build decision trees can be used following popular algorithms: ID3, CART (Classification and Regression Trees), C4.5, CHAID (Chi-Squared Automatic Interaction Detection) and QUEST (Quick, Unbiased, Efficient, Statistical Tree) [8, 9].

The decision trees method is the best option for solving classification problems, such as processing customer applications. It helps us make the best decisions based on the information we have and our best guesses. It has a number of advantages and characteristic features.

In the literature [8-10] the following advantages of using decision trees are given:

– The use of trees simplifies complex relationships between input and target variables by separating the input variables into meaningful subgroups.

– Trees are easy to understand and interpret. If a decision tree does not have many leaves, it can be understood by non-professional users. In addition, decision trees can be converted into a set of rules.

– Help to make the best decisions based on the available information.

– Decision trees can be used to process nominal input attributes as well as to process numeric ones.

– Decision trees are capable of handling datasets containing errors or missing values.

– Trees are robust to outliers.

– Decision trees are considered a non-parametric method. Trees have no assumptions about distribution and the classifier structure.

Besides, trees are able to provide high speed classification even when processing a large amount of information

Thus, thanks to the introduction of decision support systems based on decision trees in the mechanisms of processing customer applications of any institution where it is necessary to make decisions on them, there is an opportunity:

– facilitate and significantly speed up the decision-making process on applications;

– minimize the impact of the human factor and the number of errors that occur in the decision-making process;

– significantly improve the working conditions of employees of institutions by reducing the load on them;

– reduce the number of queues in institutions;

– improve the conditions for clients of institutions, by speeding up the submission and processing of their applications.

When comparing the effectiveness of decision trees and neural networks for classifying texts in information security tasks, the authors of the paper [11] note that for the task of classifying texts by quantitative features, decision trees lose in effectiveness to neural networks. But at the same time, decision trees have a significant advantage in speed, because their construction takes more than ten times less time than setting up neural networks. In addition, the answer given by decision trees is easily interpreted as a set of rules for choosing one or another class. Neural networks, on the other hand, provide only information about the degree to which objects belong to certain classes.

CONCLUSIONS

Therefore, when there is a need to process a large number of monotonous customer applications, it is appropriate to use a decision support system. The use of such systems will help to consider all applications quickly and in time, without creating additional inconvenience.

The processing of applications is reduced to their division into two classes (accept or reject). This problem belongs to the problem of binary classification. To solve such problems, it is proposed to use the method of decision trees due to the large number of its advantages. It will allow to make classification of a large number of applications in a minimum of time. It also provides a visual interpretation of the results and allows you to find out, according to the non-compliance with which criteria, a particular decision was made.

References:

1. Rasku J. Decision Support Systems / J. Rasku, P. Kuusipalo, H. Joutsijoki. // Review Working Paper from the EduMAP project. – 2020. – 21 p.

2. Гавор М. І. Розробка способу класифікації інформаційних сигналів шляхом побудови оптимального дерева рішень // Енергетика. Екологія. Людина. Зб. наукових праць ІЕЕ, КПІ імені Ігоря Сікорського – Київ: ІЕЕ, 2020. – С. 153-156

3. Matsuga O. Decision tree ensemble constructing algorithms and their comparative analysis / O. Matsuga, B. Nashylnyk // Актуальні проблеми автоматизації та інформаційних технологій. – Т. 22. – 2018. – С. 62-69

4. Федосова, І. В. Застосування систем підтримки прийняття рішень в механізмі призначення субсидій. / І. В. Федосова, Л. Д. Котихова // Наука та виробництво : зб. наукових праць / ДВНЗ «ПДТУ». – Маріуполь, 2019. – Вип. 21. – С. 178–184.

5. Kabari L. G. Decision Support System Using Decision Tree and Neural Networks / L. G. Kabari, E. O. Nwachukwu. // Computer Engineering and Intelligent Systems. – 2013. – №7. – P. 8–19.

6. Neelamegam S. Classification algorithm in Data mining: An Overview / S. Neelamegam, E. Ramaraj. // International Journal of P2P Network Trends and Technology. – 2013. – №5. – P. 1–5.

7. **Tan P.** Introduction to Data Mining (Second Edition) / **P. Tan, M. Steinbach, A. Karpatne, V. Kumar. –** Pearson, 2018. – **864 p.**

8. Maimon O. Data Mining and Knowledge Discovery Handbook / O. Maimon, L. Rokach. – Springer Science + Business Media, LLC, 2010. – 1285 p.

9. Yan-yan Song Decision tree methods: applications for classification and prediction / Yan-yan Song, Ying Lu. // Shanghai Arch Psychiatry. – 2015. – №27(2). – P. 130–135.

10. Pearl J. Using a Decision Tree to help Make Decisions [Electronic resource] / J. Pearl, T. Wardley. – 2019. – Mode of access to the resource: www.extensionaus.com.au/youngfarmerbusinessnetwork.

11. Яворська Н. Порівняння дерев рішень та нейронних мереж для класифікації текстів в задачах інформаційної безпеки / Н. Яворська, А. Яворський // Матеріали V науково-технічної конференції „Інформаційні моделі, системи та технології“, 1-2 лютого 2018 року. — Т. : ТНТУ, 2018. — С. 104. — (Секція 5. Безпека інфокомунікацій).

Котихова Л. Д., Левицька Т. О.

ЗАСТОСУВАННЯ МЕТОДУ ДЕРЕВ РІШЕНЬ ДЛЯ ЗАДАЧ КЛАСИФІКАЦІЇ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

В будь-якій галузі діяльності людини можна простежити значний вплив інформаційних технологій. Хоч люди і створюють дуже складні конструкції в різних сферах, що їх не здатен зробити жоден комп'ютер, вони значно програють при виконанні повторюваних завдань з однаковим результатом. У таких завданнях неминучий людський фактор та наявність помилок, чого можна уникнути при автоматизації процесу вирішення подібних завдань та дорученні цієї роботи комп'ютеру.

Швидкий та стрімкий розвиток інформаційних технологій призвів до появи систем підтримки прийняття рішень (СППР), що значно полегшують роботу людей у багатьох різноманітних сферах діяльності, де виникає потреба в прийнятті рішень співробітниками.

Особливого значення набуває використання систем прийняття рішень для обробки заяв людей в різних установах. Такі задачі здебільшого відносяться до задач класифікації. Для розв'язання таких задач існує декілька методів, але одним з найвідоміших є метод дерев рішень, що має ряд переваг.

Стаття присвячена аналізу застосування методу дерев рішень для завдань класифікації в системах підтримки прийняття рішень. У цій статті наведено опис об'єкта

дослідження, проведено систематизацію відомих підходів до вирішення проблеми, що розглядається, виділено та проаналізовано переваги використання методу дерев рішень у системах підтримки прийняття рішень в цілому та, зокрема, для обробки заяв людей у різних установах.

Застосування методу дерев рішень для вирішення завдань класифікації заяв допоможе значно покращити якість обслуговування клієнтів установ, зробить процес оформлення заявок та очікування рішення щодо них більш комфортним, а відповідно, сприятиме підвищенню якості життя населення.

Ключові слова. *Інтелектуальний аналіз даних, метод дерев рішень, система підтримки прийняття рішень, задача класифікації.*

Котыхова Л. Д., Левицкая Т. А.

ПРИМЕНЕНИЕ МЕТОДА ДЕРЕВЬЕВ РЕШЕНИЙ ДЛЯ ЗАДАЧ КЛАССИФИКАЦИИ В СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

В любой области деятельности человека можно проследить существенное влияние информационных технологий. Хотя люди и создают очень сложные конструкции в разных сферах, которые не способен сделать ни один компьютер, они проигрывают при выполнении повторяющихся задач с одинаковым результатом. В таких задачах неизбежен человеческий фактор и наличие ошибок, чего можно избежать при автоматизации процесса решения подобных задач и поручении этой работы компьютеру.

Быстрое развитие информационных технологий привело к появлению систем поддержки принятия решений (СППР), значительно облегчающих работу людей во многих различных сферах деятельности, где возникает потребность в принятии решений сотрудниками.

Особое значение приобретает использование систем принятия решений для обработки заявлений людей в разных учреждениях. Такие задачи в большинстве своем относятся к задачам классификации. Для их решения существует несколько методов, но одним из самых известных является метод деревьев решений.

Статья посвящена анализу применения метода деревьев решений для задач классификации в системах поддержки принятия решений. В статье приведено описание объекта исследования, проведена систематизация известных подходов к решению проблемы, рассмотрены и проанализированы преимущества использования метода деревьев решений в системах поддержки принятия решений в целом и, в частности, для обработки заявлений людей в различных учреждениях.

Применение метода деревьев решений для решения задач классификации заявлений поможет значительно улучшить качество обслуживания клиентов учреждений, сделает процесс оформления заявок и ожидания решения по ним более комфортным, а соответственно, будет способствовать повышению качества жизни населения.

Ключевые слова. *Интеллектуальный анализ данных, метод деревьев решений, система поддержки принятия решений, задача классификации.*

Стаття надійшла 22.02.2022 р.