The application of a recommendation algorithm to the selection of places and the volume of planting of forest plantations, taking into account various characteristics

Maria Korzhova[1*], *Kristina* Andreeva[1], *Daniil* Zolotarev[1], and *Vladimir* Zolnikov[2]

[1]Voronezh State Technical University, 394006 Voronezh, Russia
[2]Voronezh State University of Forestry and Technologies named after G.F. Morozov, 394087 Voronezh, Russia

Abstract. The paper considers the use of a recommendation algorithm of hierarchical clustering to optimize the choice of places and the volume of planting of forest plantations, taking into account the characteristics of the environment. The relevance of the study is related to the need to improve the efficiency of forestry in the context of climate change. The process includes data collection and analysis, on the basis of which an algorithm is developed for selecting tree species and planting volume.

1 Introduction

In the context of global climate change and increasing anthropogenic pressure on natural ecosystems, special attention is paid to the restoration and maintenance of forest resources. Optimization of reforestation processes requires an integrated approach to choosing tree planting sites and determining their volume, taking into account many factors such as climatic conditions, soil type, ecology of the area and existing infrastructure. One of the promising methods for solving such problems is the use of recommendation algorithms that allow you to take into account various characteristics of the terrain and generate recommendations that contribute to the maximum efficiency of planting forest plantations. The introduction of such technologies into forestry not only improves the accuracy of decisions, but also helps to achieve more sustainable and long-lasting results.

A combinatorial approach to restoration, combining technology, data science and local knowledge, creates a truly actionable strategy that, over time, will help not only restore forest ecosystems, but also make them more resilient to climate change and human impacts.

2 Materials and methods

Forestry tree planting is a set of activities aimed at creating sustainable forest ecosystems that provide important functions such as restoring biodiversity, preventing soil erosion and carbon

* Corresponding author: wkz@rambler.ru

sequestration. When implementing these works, it is necessary to take into account many factors in order to achieve the best results. Let's take a closer look at the main features of forestry planting. The choice of tree species, one of the key aspects of reforestation is the choice of tree species that will be adapted to specific conditions. Each tree species has its own environmental requirements, including:

• Climatic conditions: temperature ranges, duration of the growing season, the need for sunlight. For example, coniferous trees such as pine or spruce are better adapted to cold climates, while hardwoods such as oak and beech prefer warmer zones.

• Soil type: acidity, structure and nutrients. For the successful growth of trees, it is important to choose species that are adapted to the characteristics of the soil on the site, whether it is sandy, clay or stony soils.

• Humidity level: Plants requiring wetter conditions, such as willow or poplar, should be planted near reservoirs or in areas with high groundwater levels.

The terrain, the terrain of the site has a great influence on the success of planting forest crops. Different types of terrain require special approaches:

• Slope landings: Steep slopes are prone to erosion, which requires the creation of stabilizing landings. Usually, tree species with a deep root system capable of retaining soil are planted, for example, pine or oak.

• Plateaus and plains: denser tree planting can be planned in these areas, as the risk of erosion is lower and access to water and nutrients is more uniform.

Density and planting pattern. The choice of density and planting scheme has a direct impact on the viability of forest plantations. Overly dense plantings can create competition for light, water and nutrients, which negatively affects the growth of trees. On the other hand, too sparse plantings do not provide the necessary forest microclimate and reduce the protective functions of the forest.

The main approaches to planting density include:

• Mosaic planting: it is used to create heterogeneous forests where different species of trees and shrubs are combined on the same territory. This contributes to increasing the biological diversity and sustainability of forest ecosystems.

• Linear or uniform planting: used to create forest belts or protective plantings along roads and fields. This method helps to prevent wind and water erosion of soils.

Taking into account seasonality and weather conditions, an important aspect is choosing the right time for planting trees. Depending on the region, planting can be carried out in spring or autumn, when conditions are most favorable for rooting seedlings. For example [1-5]:

• Spring planting: preferred in regions with cold and humid winters, so that the trees have time to take root before the onset of summer heat.

• Autumn planting: suitable for regions with mild winters, where rooting can occur in winter, which gives seedlings an advantage before the onset of spring growth.

Protection from pests and diseases, for the successful growth of trees, it is necessary to provide measures to protect against pests and diseases. This may include:

• Regular monitoring of the condition of forest crops: using modern technologies such as satellite imagery or drones, it is possible to monitor the condition of plantations and identify early signs of disease or damage.

• Application of biological protection methods: the use of natural predators or pest antagonists to preserve the ecosystem balance.

Ecological role and restoration of biodiversity, tree planting is often aimed not only at creating forests, but also at restoring degraded ecosystems and maintaining biodiversity. To do this, it is important to:

• Use mixed planting: the combination of coniferous and deciduous trees helps to create more sustainable forest ecosystems that better cope with environmental changes.

• Conservation of natural ecosystems: when planting, it is important to take into account existing ecosystems, such as reservoirs, meadows and swamps, and not disrupt their balance.

Based on the conducted research in the field of reforestation, it was revealed that the process of choosing places and planting volume of forest plantations can be optimized using recommendation algorithms. Formally, the task of forming recommendations can be presented in the form of a mathematical record (formula 1):

$$\forall p \in P,\ o'p = argmax\ o \in O\ f(p,o) \quad (1)$$

where P is the set of plots, O is the set of possible volumes and planting sites of forest plantations that can be recommended for the site, f is a function that determines how much a certain volume and planting site satisfies the conditions of the site p. Thus, it is necessary to choose such a volume and landing site o', which belongs to the set of objects O, and at which the correspondence value for each site p ∈ P is maximum.

When using a dataset to apply a recommendation algorithm to the selection of sites and the amount of planting of forest plantations, a thorough analysis of the data is necessary. In the process of studying the literature and Internet resources, key stages of data preprocessing were identified that will help ensure the quality of recommendations:

1. Handling missing values: It is important to identify the missing data and decide how to deal with it. This may include deleting records with missing values or filling them with average or most common values to improve the completeness of data on soil characteristics, climate and other factors.

2. Conversion of categorical variables into numerical ones: Many characteristics that affect the choice of tree planting sites (for example, soil type or climatic conditions) can be presented in categorical form. They need to be converted to numeric values using encoding methods such as One-Hot Encoding or Label Encoding so that the algorithm can process this data.

3. Scaling of numerical variables: If numerical characteristics (for example, humidity level or temperature) have different ranges, then they should be scaled. This can be done through standardization or normalization, which allows the algorithm to estimate values more accurately.

4. Outlier processing: It is necessary to identify and process outliers in the data. This may include removing abnormal values from the data set on the characteristics of planned tree planting or replacing them with more typical values to avoid distorting the final recommendations.

5. Creating new features: Based on the initial data, you can create new features that may be useful for modeling. For example, you can combine data on soil types and humidity to create a new feature that reflects favorable conditions for the growth of certain types of trees.

6. Data normalization: In some cases, it may be necessary to normalize the data to improve its consistency. This is especially true when using algorithms that assume a normal distribution of data, which can improve the accuracy of recommendations.

7. Visualization and Research analysis: Data visualization and research analysis provide insight into the distribution of variables and the relationships between them. This can reveal important factors influencing the choice of landing sites, as well as detect possible problems with data quality.

Qualitative data is the basis for the successful application of forecasting algorithms. To prevent the situation of "garbage at the entrance, garbage at the exit", it is necessary to regularly check the state of the data and promptly identify problems, which will allow timely resolution of issues of pre-processing and cleaning of data. This, in turn, will increase the effectiveness of models and recommendations for the selection of sites and the volume of planting of forest plantations, taking into account various characteristics [6-10].

3 Results and analysis

Hierarchical clustering is a significant data preprocessing method that can be used before implementing a recommendation algorithm to determine the location and volume of planting of forest stands. This approach allows you to analyze various characteristics that affect the growth of trees, increasing the efficiency of the process.

Collaborative filtration applied in this context helps to predict preferences for planting forest plantations based on historical data on climatic conditions, soil types, water availability and other important factors. Hierarchical clustering divides this data into groups, or clusters, based on common characteristics. Each piece of land with certain conditions starts with a separate cluster, after which they are combined, which forms a hierarchy of data.

Using the hierarchical clustering method before filtering together is extremely useful, because it allows you to reduce the dimensionality of the data by combining plots with similar conditions. With large amounts of data consisting of many variables, this approach facilitates analysis and decision-making. Clustering helps to collect plots with similar conditions in one cluster, which, in turn, can improve the quality of tree planting recommendations, since plots with similar characteristics will receive advice that matches their conditions.

In addition, hierarchical clustering facilitates the processing of missing data. If there are empty values in the characteristics of the plots, clusters can be used to fill them. Information about clusters will allow the site to be assigned to more complete groups, which will provide more accurate recommendations based on similarities with other sites in the cluster.

Thus, hierarchical clustering turns out to be a valuable tool in the data preparation process for applying a recommendation algorithm to the selection of sites and the volume of planting of forest plantations. It improves the quality of recommendations, reduces the dimensionality of data, and effectively handles missing values.

When developing recommendation systems for selecting tree planting sites, the hierarchical clustering method will be used, especially when processing large amounts of data including millions of records. With an increase in the amount of data, the speed of searching for optimal landing sites becomes significantly time- and resource-intensive. A preliminary search for cluster centers with further refinement within the selected cluster is one way to speed up the recommendation process. This initiating hierarchical clustering may take a lot of time at the creation stage, but it will be updated less frequently, for example, once a day, as new data becomes available, rather than with each user request. As a result of hierarchical clustering, a dendrogram is created that displays the nesting of clusters, starting with small ones and gradually moving to larger ones, which clearly demonstrates the complex relationships in the data.

The tree search is performed in the opposite direction from top to bottom, by sequentially selecting the closest cluster centers. The gain in search speed depends on the number of hierarchical levels. Let's illustrate this with a simple example. The path has five levels of clustering with branching at each level into three branches. Each last cluster contains 1000 records. Total records N=1000*35=243,000. With direct filtering, you need to make 243,000 comparisons and select the nearest instance. When descending the clustering tree, you need to make 1000+3*5 =1015 comparisons. The resulting gain using clustering is obvious.

To form recommendations, it is possible to use a test dataset in a tabular file. As mentioned earlier, it is necessary to clean up the data, for example, using the Pandas library

The result of the execution is a Plot-volume matrix. After this matrix has been built, it is necessary to calculate two new matrices based on it with coefficients of similarity (similarity, proximity) for plots and for planting volumes.

In this paper, the cosine distance between vectors is used as a proximity metric. To calculate the cosine distance, we use the sklearn library. Cosine distance helps

recommendation systems to offer more accurate and personalized results, improving the overall user experience.

To assess the quality of the prediction, we use the RMSE metric (Root Mean Square Error). The result of the execution is recorded in a text file and contains information about the error that occurs during the recommendation process. The higher the RMSE value, the worse the quality of the recommendations [11-13].

4 Conclusions

Hierarchical clustering is an important tool for the implementation of systems that recommend places and volumes of planting of forest plantations. It allows you to group land plots and tree species into clusters, which contributes to a more accurate prediction of optimal planting sites based on various characteristics such as soil type, climatic conditions and water availability.

In general, the successful application of the recommendation algorithm for the selection of sites and the volume of planting of forest stands depends on the qualitative optimization of hierarchical clustering. The right choice of optimization methods can significantly improve the efficiency of forest plantation planning and improve the ecosystem in the region. Optimization of hierarchical clustering includes aspects such as reducing the dimensionality of data, choosing an appropriate clustering algorithm, and using parallel computing. These methods can significantly speed up the process of determining the most suitable areas for reforestation and improve the quality of recommendations, for example, in terms of volume and type of planting material.

References

1. B.Ya. Karastelev, D. S. Rogulin, N. S. Rogulin, Problems of management theory and practice **2**, 142-152 (2020)
2. A.I. Pyzhev, R.V. Gordeev, E.A. Vaganov, Sustainability **13(1)**, 86 (2020)
3. I. Avdeeva et al., IOP Conference Series: Earth and Environmental Science. – IOP Publishing **595(1)**, 012051 (2020)
4. D. Dobrynin et al., Ambio **50(12)**, 2138-2152 (2021)
5. A. Raihan, A. Tuspekova, Journal of Environmental Science and Economics **1(4)**, 1-11 (2022)
6. M. Yu. Sheresheva, *Information systems of Russian companies in industries with natural monopoly.*Theoretical and practical problems of modern economic reality. Materials of the scientific conference **5** (2017)
7. N. E. Antonova, N. V. Lomakina, Institutional innovations for the development of the East of Russia: effects of implementation in the resource region (2020).
8. N. V. Savvin, K. A. Andreeva, N. S. Pechkurov, R. S. Cheremisin, Student and science **3(22)**, 24-27 (2022) EDN KECEZQ (in Russian)
9. Development of artificial intelligence technologies in Russia: goals and reality [Electronic resource] // Access mode: https://carnegieendowment.org/2020/07/07/ru-pub-82173 - free (in Russian).
10. N. V. Menshikova, I. V. Portnov, I. E. Nikolaev, Academy **6(9)**, 20-22 (2016)
11. A. I. Pisareva, *ALGORITHMS OF RECOMMENDATION SYSTEMS: PROBLEMS AND IMPLEMENTATION*. Problems and prospects for the development of the scientific and technological space, pp. 482-485 (2020)

12. S. A. Amelkin, Proceedings **14**, 288-291 (2012)
13. C.I. Chalco, R.B. Chasi, R.H. Ortiz, *Hierarchical Clustering for Collaborative Filtering Recommender Systems*. In: Ahram, T. (eds) Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2018. Advances in Intelligent Systems and Computing, **787**. Springer, Cham (2019) https://doi.org/10.1007/978-3-319-94229-2_34