Decision Tree Methods in Grain Yield Forecasting

Peter Grabusts
Rezekne Academy of Technologies
Institute of Engineering
Rezekne, Latvia
peteris.grabusts@rta.lv

Oleg Uzhga-Rebrov
Rezekne Academy of Technologies
Institute of Engineering
Rezekne, Latvia
rebrovs@tvnet.lv

Lasma Prizevoite
Rezekne Academy of Technologies
Faculty of Engineering
Rezekne, Latvia
lasma.prizevoite@gmail.com

Inta Kotane
Rezekne Academy of Technologies
Faculty of Economics and Management
Rezekne, Latvia
inta.kotane@rta.lv

Abstract. **Traditional methods for forecasting yield, which rely on human judgment, often fall short of providing accurate and reliable forecasts. The application of artificial intelligence (AI) methods for predicting grain harvest is becoming increasingly relevant to balance performance indicators of companies in the grain-growing industry and forecast future results. It is important to consider the specific operations of companies in the industry and the factors influencing the harvest when using such methods, as these are essential for future decision-making. The main goal of the study is to explore the use of decision analysis methods in forecasting the yield of companies in the grain-growing industry. An analytical study has been conducted on the potential of using AI methods, including the analysis of decision tree-building methods and their application possibilities. In a practical study, a decision tree is constructed using CHAID, and the impact of various factors on decision-making quality in the grain-growing industry is analyzed. Subsequently, neural networks are used to predict potential yield based on the companies' historical data from previous periods.**

Keywords: CHAID, decision analysis, decision tree, grain yield, neural networks.

I. Introduction

The topicality of the topic is because the continuous changes and challenges of the grain-growing industry, which include the influence of various factors on grain harvest volumes, require grain-growing companies to adapt to new technologies, including the opportunities offered by AI, to maintain competitiveness and promote sustainable development.

AI can be used in the agricultural sector to improve production efficiency, resource management, and decision-making. In recent years, the world has seen rapid advances in agricultural technology, changing and improving farming practices. These innovations are becoming increasingly important as global challenges such as climate change, population growth, and resource scarcity threaten the sustainability of the food system. The introduction of AI solves many problems and helps to alleviate several disadvantages of traditional agriculture [1]-[6].

It is neural networks (NN) that are increasingly used in research for the needs of the agricultural industry [7], the learning ability and accuracy of NN in various stages of agriculture achieve remarkable results [8], thus they are used in production impact prediction and other aspects of agriculture based on a wide range of a range of independent variables, thus optimizing the storage and transportation processes and allowing to predict the incurred costs depending on the chosen direction of management [9]. The spectrum of applications of NN is very wide, they are used to support agricultural production, making it more efficient and providing the highest possible quality products [10].

To achieve more effective business performance, industry specialists and managers should use the advantages provided by AI by choosing suitable AI technologies and adapting them to the specifics of the industry and the company, thus improving the accuracy of economic activity, decision-making, customer experience, optimization of resources, and as a result, improving the overall results of the company or efficiency and competitiveness indicators of the organization.

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol2.8023

Cereal farming is one of the leading agricultural sectors, where all-season farmers must face the challenge of effectively planning and implementing activities to increase yield and quality.

Yield forecasting is a critical factor in grain-growing enterprises, as it affects resource use decisions, that is, the use of agricultural land and cash, as well as the quality and quantity of harvest. Traditional yield forecasting methods are often unable to provide sufficiently accurate and reliable forecasts, as it is impossible to predict individual factors affecting agriculture based on them. AI application provides new crop forecasting opportunities in the grain-growing industry [11],[12].

In general, crop forecasting modeling based on AI methods and data analysis is an indispensable tool for companies in the grain-growing industry, ensuring greater efficiency, sustainable production, and competitiveness in the market. This allows companies to respond to the continuous changes determined by both natural conditions and the market situation.

The study aims to investigate the application of decision analysis methods in forecasting the yield of companies in the grain-growing industry. Decision tree-building methods and their application possibilities are analyzed. Forecasting of potential yield is performed based on historical company data for previous periods.

II. Materials and Methods

A. Grain Yield Factors

Cereal farming is defined as the basic branch of agricultural production – the cultivation of cereals to supply the food industry with raw materials and livestock with fodder. Cereal cultivation is one of the most popular agricultural sectors. Every new season is like a challenge for farmers when they must plan actions to increase yield and quality. Tillage, sowing, fertilizing, plant protection - everything is planned in favor of the harvest, considering account the farm's financial condition, so that the result is economically efficient.

Cereal harvest volumes may depend on various factors influencing them; therefore, it is important to study their parameters and influence indicators (factors influencing the yield of winter cereal crops will not be considered within the scope of the work).

It can be concluded that such factors as grain weight, grain sowing depth, seed quality, soil quality, sowing rate, air temperature, rainfall, and type of tillage can affect grain yield, and for each of the factors, an index of influence on grain yield is defined.

The authors analyze the following factors influencing the grain yield and their impact indicators (see Table 1). The code system is provided by the authors.

As can be seen in Table 1, an important factor that affects the amounts of grain harvest is the seed germination of the grain variety used. Early growth and vigor of many cereal cultivars have been determined to be affected by seed size and weight. Based on the study [13], it is determined that the optimal sowing depth is 6 cm, which ensures the durability of seedlings and good results of crop germination.

Harvest volumes are influenced by the selected grain seed quality. The main indicator of seed quality is the protein content in the grain, where the E quality class is designated as high, with a protein content in the grain >14.5%, the A quality class is medium, with a protein content in the grain 14%, as well as the B quality class is low, with a protein content in the grain 12 - 13%.

An important factor is the quality of the soil, an important indicator characterizing the quality of the soil is the pH level, which is relatively easy for the farmer to determine by soil analysis, which for cereals in general is on average 6.0 to 6.5 pHKCl.

Table 1 Compendium of Factors Affecting Grain Yield

N	Factor	Affects	Impact score	Co-de
1	Germination (grain weight per thousand grains)	Favorable	51 grams and over	1
		Medium	26 – 50 grams	2
		Unfavorable	1 – 25 grams	3
2	Germination (grain sowing depth)	Favorable	60 – 69 mm	1
		Medium	70 – 83 mm; 23 – 59 mm	2
		Unfavorable	84 mm and more; 22 mm and less	3
3	Grain seed quality (protein content)	Favorable	14.5% and more	1
		Medium	13.1 – 14.4%	2
		Unfavorable	13% and less	3
4	Soil quality (pHKCl level)	Favorable	6 – 6.5	1
		Unfavorable	5.9 and below 6.6 and above	2
5	Sowing rate (amount of viable seeds per m2)	Favorable	400 – 500 and more	1
		Medium	300 – 399	2
		Unfavorable	299 and less	3
6	Temperature	Favorable	20 – 25 °C	1
		Medium	26 – 31 °C; 7 – 25 °C	2
		Unfavorable	32 °C and above; 6 °C and below	3
7	Amount of precipitation per year	Favorable	50-100 cm	1
		Unfavorable	49 cm and less 101 cm and above	2
8	Type of soil treatment	Traditional tillage	-	1
		Minimum tillage	-	2
		Direct sowing	-	3

The seeding rate can affect the size of the crop, thus its quality and yield volumes. Based on the data collected in Table 1, it can be concluded that the highest harvest volumes can be achieved if the sowing is on average 400-500 viable seeds per m^2 .

The growth process of cereals is significantly affected by the temperature range. The optimum temperature is 20-25°C, which means that germination is more efficient in this temperature range. Each 1°C increase in temperature above the average temperature of 23°C reduces grain yield by about 10% [14], while temperatures above 32°C negatively affect cereal growth, and temperatures below 6°C for long periods are critical [15].

Rainfall is an important factor in the grain-growing industry. If there is insufficient rainfall, cereals cannot grow, and yields may be limited. Most cereals require between 50 and 100 cm of rainfall per year [16]. Moisture or drought stress causes about 30-70% loss of cereal productivity during the crop growing period [17].

As can be seen in Table 1, three types of tillage are distinguished with different characteristic indicators. Traditional tillage refers to plowing, which includes turning the soil, which ensures the availability of nutrient elements in the entire layer of the arable layer, limits weeds, and facilitates the easier execution of other technological operations. Minimum tillage is a method where the soil is not turned over. The maximum depth of cultivation is no deeper than 10 cm and/or the percentage of plant residues left on the soil surface is determined in percent, usually 30%, applying the mentioned method reduces the risk of erosion and crust formation because straw residues remain on the soil surface [9]. Direct sowing, on the other hand, is a method where the seed is placed in the soil without cultivating the previous crop, mainly used in dry regions.

Considering all the invoices affecting the yield, the authors selected the following average yield intervals and assigned them the corresponding codes to be used as classifiers in decision trees and NN (see Table 2). We define 5 classes that describe the quality of the harvest.

TABLE 2 AVERAGE YIELD INTERVALS AND THEIR ASSIGNED CODES

Average yield interval (t/ha)	Affects	Code
1.0 - 1.9	Very unfavorable	1
2.0 - 2.9	Unfavorable	2
3.0 - 4.0	Medium	3
4.1 – 5.1	Favorable	4
5.2 - 6.5	Very favorable	5

Grain yield forecasting is an important aspect of agriculture that helps farmers make informed decisions about crops grown [18], which is essential for resource optimization and investment planning for sustainable production [19].

B. Decision Tree Construction Methods

Decision trees are one of the most effective data mining tools that allow you to solve classification and regression problems. They are hierarchical tree structures consisting of decision rules of the form “If ... Then ...”. The rules are automatically generated during the learning process on the training set and, since they are formulated almost in natural language, decision trees as analytical models are more verbalizable and interpretable than NN.

The decision tree is a method of representing decision rules in a hierarchical structure consisting of two types of elements - nodes and leaves. The nodes contain decision rules and check the compliance of examples with this rule using any attribute of the training set.

Then the rule is applied again to each subset and the procedure is repeated recursively until a certain condition for stopping the algorithm is reached. As a result, the last node is not checked or split and is declared a leaf. The worksheet determines the solution for each example included in it. For a classification tree, this is the class associated with the node, and for a regression tree, this is the modal interval of the target variable corresponding to the leaf.

Let a training set S be given, containing n examples, for each of which a class label is given $C_i(i=1..k)$ and m attributes $A_j(j=1..m)$, which are assumed to determine whether an object belongs to a particular class. Then three cases are possible:

1. All examples of the set S have the same class label C_i (that is, all training examples belong to only one class). Training in this case does not make sense, since all the examples presented to the model will be of the same class, which will “learn” to recognize the model. The decision tree itself in this case will be a leaf associated with the class C_i. The practical use of such a tree is pointless since it will assign any new object only to this class.
2. The set S does not contain examples at all, i.e. is the empty set. In this case, a leaf will also be created for it (applying a rule to create a node to an empty set is pointless), the class of which will be selected from another set (for example, the class that occurs most often in the parent set).
3. The set S contains training examples of all classes C_k. In this case, it is necessary to split the set S into subsets associated with classes. To do this, select one of the attributes A_j of the set S which contains two or more unique values $(a_1,a_2,...,a_p)$, where p is the number of unique values of the attribute. The set S is then split into p subsets $(S_1,S_2,...,S_p)$, each of which includes examples containing the corresponding attribute value. Then the next attribute is selected, and the partition is repeated. This procedure will be repeated recursively until all examples in the resulting subsets are of the same class.

The procedure described above underlies many modern algorithms for constructing decision trees. When using this technique, the construction of a decision tree will occur from top to bottom (from the root node to the leaves).

Currently, a significant number of decision tree learning algorithms have been developed: ID3 (Iterative Dichotomizer 3), CART (Classification and Regression Tree), C4.5, C5.0, NewId, ITrule, CHAID (Chi-square automatic interaction detection), CN2, etc. [20-21].

CHAID determines the relationship between a response variable and others, so you can forecast how to have the biggest impact. The CHAID algorithm splits nodes to produce chi-square values. A chi-square value is the difference between a standard and the results observed in your data. The maximum chi-square value is the most statistically significant result in the CHAID decision tree. It is the strongest relationship between two variables of the found chi-square values.

Advantages of the CHAID algorithm:

- fast learning process.
- generation of rules in areas where it is difficult for an expert to formalize his knowledge.
- extracting rules in natural language.
- intuitive classification model.
- high prediction accuracy, comparable to other data analysis methods (statistics, NN).
- construction of nonparametric models.

III. Results and Discussion

To characterize the enterprises of the grain-growing sector and their performance indicators, four agricultural enterprises whose main activity is grain cultivation were selected. The study used real data and the initial data contained the indicators of the factors influencing the harvest and the harvest volumes in the period from 2011 to 2023.

The research methodology is based on the following sequence of actions:

1. A decision tree is created based on the initial data.
2. Based on these data, an NN is constructed and trained.
3. NN testing is carried out to be able to predict the yield at certain factor indicators.

The resulting decision tree is shown in Figure 1.

As can be seen in Figure 1, the CHAID decision tree is divided into 12 nodes, and for each of the factors, a chi-square value is indicated, which indicates the difference between the expected result and the actual data. The root node is the yield interval, which is the starting point in the decision tree and on which all other data depends.

The highest chi-square value is observed for the factor “precipitation”, which means that this is the most statistically significant result in the created decision tree. The decision tree is divided into further nodes, after which the most significant factor that has the greatest influence on the dependent factor becomes the new variable, thus indicating which variables are most effective for this distribution of data.

Based on the collected indicators, because of a result of the CHAID decision tree analysis, the importance of factors is determined: temperature, grain weight, precipitation, sowing depth, and seeding rate.

The CHAID decision tree model summarizes the influence and importance of the factors but requires the addition of NN analysis to be able to predict grain yields. A Multilayer Perceptron (MLP) was chosen for training. Also, in the NN, the yield interval is defined as a target that depends on the other factors influencing the yield. The training quality of the created NN model is defined as 71.2%, which means that the model is relatively accurate. As a result of NN training, the importance of factors is determined: temperature, precipitation, sowing depth, sowing rate, grain weight, seed quality, and soil quality.

The important factors obtained by training decision trees and NN differ because in both cases different training algorithms are used, and as a result, the factors may have different effects. Each algorithm uses different characteristics and relationships inherent in the data.

Based on the data trained by the NN, companies in the grain-growing industry can make forecasts of the next year's harvest, considering the obtained reliability index, which would provide an opportunity to identify and eliminate possible yield reductions or other adverse factors affecting the harvest, which generally affect the company's operation.

For example, based on the results of the created forecasting model on the possible indicators of the 2024 grain harvest with the changed value of the “precipitation” factor from favorable to unfavorable, it can be concluded that considering the known possible factors affecting the harvest, the 2024 harvest with a 77% reliability index also, in this case, will be very unfavorable.

To increase the prediction reliability of the created model, a much larger and more diverse amount of data would be needed (60 records were used in our study), which would allow the model to learn from more versatile data, thus significantly improving the prediction efficiency.

IV. Conclusions

Forecasting in the grain-growing industry contributes to sustainable agriculture, as more accurate forecasts help reduce the use of unproductive resources, and by using crop forecasting modeling, grain companies can more effectively adapt to fluctuations in market demand, ensuring that the yield is in line with demand, but does not exceed it, preventing potential losses.

Companies in the grain-growing sector need to gather data on the factors influencing grain yield and production volumes annually. By utilizing a pre-existing forecasting model, they can then predict the grain yield for the following year. This enables the company to adjust to evolving conditions, optimize resources, and implement necessary measures to enhance potential yield volumes.

The decision tree model could be enhanced by incorporating other factors that affect grain yield, enabling a more comprehensive evaluation of potential future grain yield.

Conceptually, the solution developed in this study is not groundbreaking, but its implementation in cereal production companies offers a glimpse into the use of modern technologies in practical operations. It can be inferred that predicting grain yield is a crucial element of agriculture that aids in resource optimization. CHAID decision trees and NN are valuable tools in this area, offering more precise predictions and aiding in decision-making for agricultural processes.

The authors conclude that the developed solution is viable and enables yield prediction based on the developed proposed yield impact factors.

The future direction of research will be related to the introduction of additional factors that will make the model work more accurately.

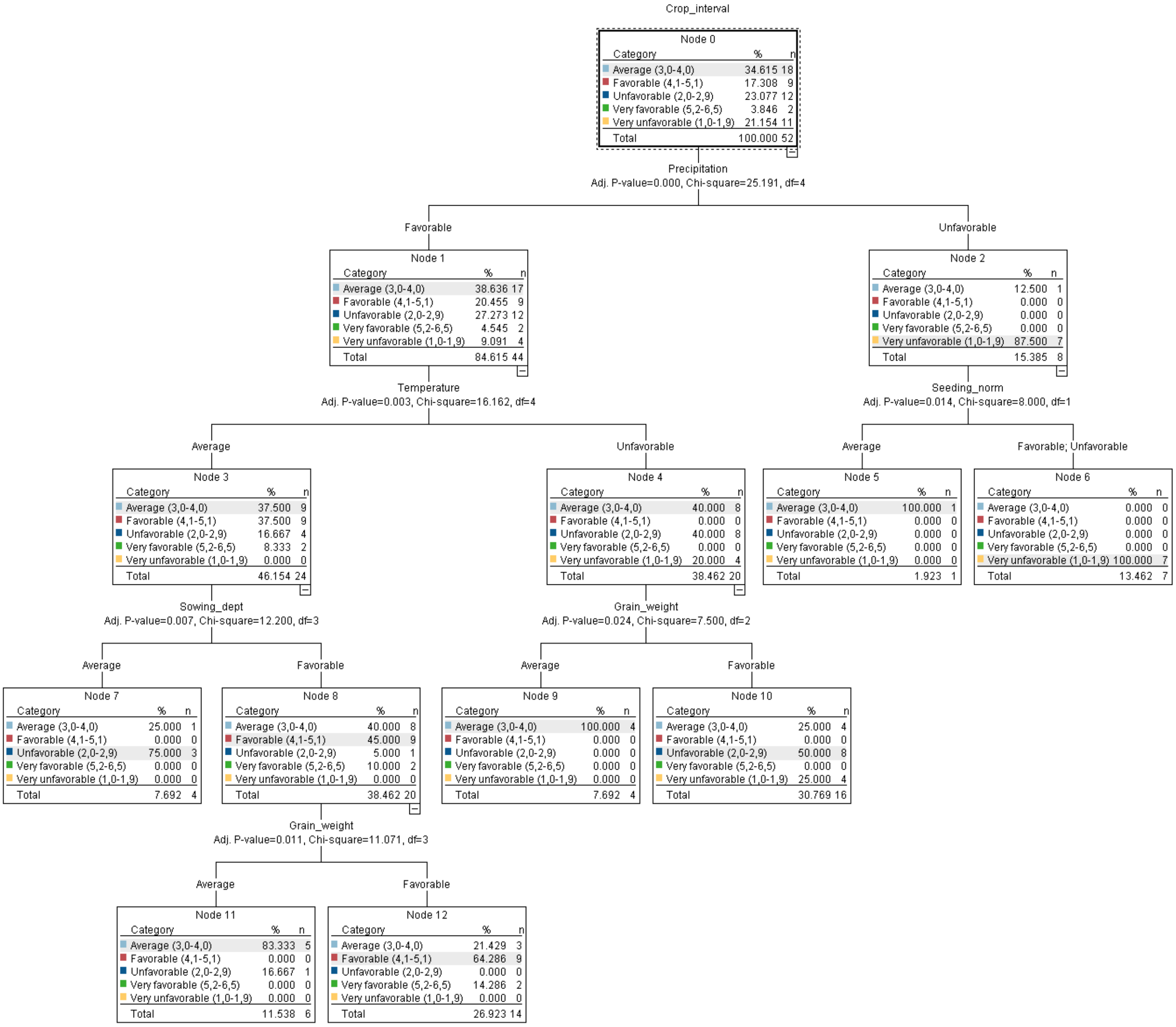
Crop_interval
Node 0
Category % n
Average (3,0-4,0) 34.615 18
Favorable (4,1-5,1) 17.308 9
Unfavorable (2,0-2,9) 23.077 12
Very favorable (5,2-6,5) 3.846 2
Very unfavorable (1,0-1,9) 21.154 11
Total 100.000 52
Precipitation
Adj. P-value=0.000, Chi-square=25.191, df=4
Favorable
Node 1
Average (3,0-4,0) 38.636 17
Favorable (4,1-5,1) 20.455 9
Unfavorable (2,0-2,9) 27.273 12
Very favorable (5,2-6,5) 4.545 2
Very unfavorable (1,0-1,9) 9.091 4
Total 84.615 44
Unfavorable
Node 2
Average (3,0-4,0) 12.500 1
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 0.000 0
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 87.500 7
Total 15.385 8
Temperature
Adj. P-value=0.003, Chi-square=16.162, df=4
Seeding_norm
Adj. P-value=0.014, Chi-square=8.000, df=1
Average
Node 3
Average (3,0-4,0) 37.500 9
Favorable (4,1-5,1) 37.500 9
Unfavorable (2,0-2,9) 16.667 4
Very favorable (5,2-6,5) 8.333 2
Very unfavorable (1,0-1,9) 0.000 0
Total 46.154 24
Unfavorable
Node 4
Average (3,0-4,0) 40.000 8
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 40.000 8
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 20.000 4
Total 38.462 20
Average
Node 5
Average (3,0-4,0) 100.000 1
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 0.000 0
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 0.000 0
Total 1.923 1
Favorable; Unfavorable
Node 6
Average (3,0-4,0) 0.000 0
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 0.000 0
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 100.000 7
Total 13.462 7
Sowing_dept
Adj. P-value=0.007, Chi-square=12.200, df=3
Grain_weight
Adj. P-value=0.024, Chi-square=7.500, df=2
Average
Node 7
Average (3,0-4,0) 25.000 1
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 75.000 3
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 0.000 0
Total 7.692 4
Favorable
Node 8
Average (3,0-4,0) 40.000 8
Favorable (4,1-5,1) 45.000 9
Unfavorable (2,0-2,9) 5.000 1
Very favorable (5,2-6,5) 10.000 2
Very unfavorable (1,0-1,9) 0.000 0
Total 38.462 20
Average
Node 9
Average (3,0-4,0) 100.000 4
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 0.000 0
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 0.000 0
Total 7.692 4
Favorable
Node 10
Average (3,0-4,0) 25.000 4
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 50.000 8
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 25.000 4
Total 30.769 16
Grain_weight
Adj. P-value=0.011, Chi-square=11.071, df=3
Average
Node 11
Average (3,0-4,0) 83.333 5
Favorable (4,1-5,1) 0.000 0
Unfavorable (2,0-2,9) 16.667 1
Very favorable (5,2-6,5) 0.000 0
Very unfavorable (1,0-1,9) 0.000 0
Total 11.538 6
Favorable
Node 12
Average (3,0-4,0) 21.429 3
Favorable (4,1-5,1) 64.286 9
Unfavorable (2,0-2,9) 0.000 0
Very favorable (5,2-6,5) 14.286 2
Very unfavorable (1,0-1,9) 0.000 0
Total 26.923 14

Fig. 1. Decision tree of CHAID performance result.

References

[1] "AI in Agriculture — The Future of Farming," [Online] Available: https://intellias.com/artificial-intelligence-in-agriculture/

[2] M. Javaid, A. Haleem, I.H. Khan, and R. Suman, "Understanding the potential applications of Artificial Intelligence in Agriculture Sector," Advanced Agrochem, Vol. 2, Issue 1, 2023. https://doi.org/10.1016/j.aac.2022.10.001

[3] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," Artificial Intelligence in Agriculture, Vol. 4. p.58-73, 2020. https://doi.org/10.1016/j.aiia.2020.04.002

[4] J. Jung, M. Maeda, A. Chang, M. Bhandari, A. Ashapure, and J.L. Bowles, "The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems," Current Opinion in Biotechnology, Vol. 70. p.15-22, 2021. https://doi.org/10.1016/j.copbio.2020.09.003

[5] A. Subeesh and C.R. Mehta, "Automation and digitization of agriculture using artificial intelligence and internet of things," Artificial Intelligence in Agriculture, Vol. 5, p.278-291, 2021. https://doi.org/10.1016/j.aiia.2021.11.004

[6] D.I. Patricio and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," Computers and Electronics in Agriculture, Vol. 153, p.69-81, 2018. https://doi.org/10.1016/j.compag.2018.08.001

[7] Y. Amapatzidis, V. Partel, and L. Costa, "Agroview: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence," Computers and Electronics in Agriculture, Vol. 174, 2020. https://doi.org/10.1016/j.compag.2020.105457

[8] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S.D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," Artificial Intelligence in the Life Sciences, Vol. 1., 2021. https://doi.org/10.1016/j.ailsci.2021.100010

[9] S. Kujawa and G. Niedbala, Artificial Neural Networks in Agriculture. MDPI, 284 p., 2021. https://doi.org/10.3390/books978-3-0365-1579-3

[10] G. Niedbala and S. Kujawa, "Artificial Neural Networks in Agriculture," Agriculture, Vol. 11, Issue 6, 2021. http://dx.doi.org/10.3390/agriculture11060497

[11] K.N. Patel, S. Raina, and S. Gupta, Artificial Intelligence and its Models. JASC: Journal of Applied Science and Computations, 2020.

[12] A. Sharma, M. Georgi, M. Tregubenko, A, Tselykh, and A. Tselykh, "Enabling smart agriculture by implementing artificial intelligence and embedded sensing," Computers & Industrial

Engineering, Vol. 165, 2022. https://doi.org/10.1016/j.cie.2022.107936 [Accessed: Feb. 1, 2024].

[13] J.J. Blake, J.H. Spink, and C. Dyer, "Factors affecting cereal establishment and its prediction," 2003. [Online] Available: https://projectblue.blob.core.windows.net/media/Default/Research%20Papers/Cereals%20and%20Oilseed/rr51_complete_final_report.pdf [Accessed: Feb. 1, 2024].

[14] S. Narayanan, "Effects of high temperature stress and traits associated with tolerance in wheat," Open Access Journal of Science, Vol.2, Issue 3, p.177-186, 2018. https://doi.org/10.15406/oajs.2018.02.00067

[15] M. Gammans, P. Merel, and A. Ortiz-Bobea, "Negative impacts of climate change on cereal yields: statistical evidence from France," Environmental Research Letters, Vol. 12, Issue 5, 2017. DOI 10.1088/1748-9326/aa6b0c

[16] NICHE Agriculture, "How Rainfall Affects Crop Health," [Online] Available: , from https://www.nicheagriculture.com/how-rainfall-affects-crop-health/ [Accessed: Feb. 1, 2024].

[17] T.N. Liliane and M.S. Charles, "Factors Affecting Yield of Crops," Agronomy - Climate Change & Food Security, 2019. DOI: 10.5772/intechopen.90672

[18] JavaTpoint, "Crop Yield Prediction Using Machine Learning," [Online] Available: https://www.javatpoint.com/crop-yield-prediction-using-machine-learning [Accessed: Feb. 1, 2024].

[19] A. Sreerama and B.M. Sagar, "A Machine Learning Approach to Crop Yield Prediction," International Research Journal of Engineering and Technology (IRJET), Vol. 07, Issue 05, 2020. [Online] Available: https://www.irjet.net/archives/V7/i5/IRJET-V7I51246.pdf [Accessed: Feb. 1, 2024].

[20] S. Murthy, "Automatic construction of decision trees from data: A Multi-disciplinary survey," Data Mining and Knowledge Discovery, Volume 2, pages 345–389, 1998.

[21] L. Rokach and O. Maimon, Data Mining with Decision Trees: Theory and Applications (2nd Edition), (Series in Machine Perception and Artificial Intelligence, vol. 81, 2015.