The Journal of Animal & Plant Sciences, 28(4): 2018, Page: 1085-1093
ISSN: 1018-7081

AN ANALYSIS OF FACTORS AFFECTING YIELD, OIL PRODUCTION RATE AND PLANT HEIGHT IN SUNFLOWERS USING SELECTED DATA MINING ALGORITHMS

S. Celik[1], E. Boydak[2] and Rıdvan Fırat[2]

[1]Department of Animal Science, Faculty of Agriculture, University of Bingol, Turkey.
[2]Department of Field Crops, Faculty of Agriculture, University of Bingol, Turkey.
Corresponding Author E-mail: senolcelik@bingol.edu.tr

ABSTRACT

In this study, yield per decare, plant length and oil production rate of sunflowers grown in the city of Bingöl, Turkey**,** and the elements affecting the plant length were examined using several data mining methods: Chi-square Automatic Interaction Detection (CHAID), Exhaustive CHAID, Classification and Regression Trees (CART), and artificial neural networks like Multi-Layer Perceptrons (MLP) and Radial Basis Function networks (RBF). Yield per decare was affected by oil production, the 1000 grain ratio, and the kernel percentage. Oil production rate was affected by the kernel percentage, plant length, 1000 grain ratio and yield per decare. Plant length was affected by the kernel percentage. These effects were significant. According to the Exhaustive CHAID algorithm used to predict yield per decare, the Pearson correlation coefficient (r), coefficient of determination (R^2), adjusted coefficient of determination (Adj. R^2), RRMSE, RAE and SD ratio were recorded as 0.915, 0.837, 0.825, 0.00025, 0.060 and 0.403, respectively. The largest average yield per decare (297.643 kg) was obtained from the subgroup consisting of plants whose oil production rate was ≤ 28.040 and grain ratio was > 83.800. According to the CART algorithm used to predict oil production rate (%), r, R^2, Adj. R^2, RRMSE, RAE and SD ratio, the results were calculated to be 0.988, 0.976, 0.975, 0.00045, 0.0021 and 0.155, respectively. The largest average oil production rate was obtained from the sub-group consisting of plants whose kernel percentage was > 62.500, whose plant length was >127.500 cm and whose kernel percentage was > 72.250. According to the Exhaustive CHAID algorithm used to predict plant length, r, R^2, Adj., R^2, RRMSE, RAE and SD ratio, these were calculated to be 0.942, 0.887, 0.883, 0.00021, 0.029 and 0.335, respectively. The longest average plant length (166.600 cm) was obtained from the subgroup consisting of plants where 69.500 ≤ kernel percentage < 70.800.
As a result, the use of Exhaustive CHAID and CART algorithms can be employed in field crops studies to predict some plant characteristics of sunflowers.

Keywords: Data mining, sunflower, plant characteristics.

INTRODUCTION

The production of sunflowers is greatly influenced by the proper hybrid selection. When choosing a hybrid, it should be focused on potential yield and other parameters contributing to the achievement of higher levels of production, such as stalk strength, disease resistance, oil content and maturity. With the increase of plant population per unit area of sunflower crop, decreased head diameters, numbers and achene weight per head, along with a higher plant density above a certain threshold has had a negative effect on the achene's yield (Mojiri and Arzani, 2003). Plant density is invariably linked with yield. With more plants, higher production is expected up to a certain limit (Bertoria *et al*., 1998).

The sunflower (*Helianthus annus L.)* is an annual summer plant of the compositae family and one of 67 species within the Helianthus genus. It is divided into two types: oil producing and edible. The oil types' seeds are small and they have high rates of oil production. According to archeological findings, the sunflower was first planted and consumed by Native American tribes (Kaya *et al*, 2005).

Among the plants with oily seeds planted in Turkey, the sunflower is the largest in terms of plant area and production. Sunflowers have become the most important oil plant in Turkey, with the largest herbal oil consumption and the highest oil production rate (Anonymous, 2014).

According to data from the Turkish Statistical Institute, sunflowers have been planted over an area of 5524 hectares; 1480 tons have been produced and the average efficiency reached about 269 kg/da by 2014. Regarding edible sunflower, while it has been planted over an area of 1490 decares, its production has been 157,000 tons, and its average efficiency has been about 152 kg/da (Anonymous, 2014).

The efficiency obtained from plant growing is a result of mutual interaction of the genotype and the environment, where the environmental circumstances consist of factors like climate, soil structure, and growing techniques (crop rotation, crop time, watering, etc.). As in the case of other agriculturally significant plants, it is important to know the physiological and morphological

features and the genotype of the preferred species, along with required agricultural operations (such as growing techniques, precautions for plant protection) to increase the efficiency in a unit area for growing sunflowers (Vasudevan *et al*., 1997).

Sunflower species that show differences in physiological and morphological features can react differently to different growing circumstances. In that sense, species-adaptation studies are highly significant to determine which species have high seed and oil efficiency, are resistant to illness and pests, and are suitable to a region's circumstances.

Oil has an important role for human nutrition and is mainly supplied worldwide from oil plants. However, an increasing gap in the field of food has been observed since herbal oils have recently become raw material for energy sectors such as biodiesel, alongside those for nutritional purposes (Poyraz, 2012).

Herbal oils are significant in terms of their contributions to human health, and have high nutritional value due to features like low saturated oil ratios, free fatty acids required for cell structure, and an ability to absorb oil-soluble vitamins like vitamin A, D, E and K (Göksu, 2007).

Sunflower oil is rich in linoleic acid (50 to 51%). Linoleic acid, due to its desiccant properties, has an important use in the dyeing industry. The oil cake obtained after degreasing it by 40 to 45%, contains 30 to 40% protein and is valuable fodder. Because of these properties, its production has increased significantly both in Turkey and worldwide (Arıoğlu, 1999).

The aim of this study was to define the best algorithm – in terms of its predictive performance – from CART (Classification and Regression Trees), CHAID (Chi-square Automatic Interaction Detection), Exhaustive CHAID and ANN (Artificial Neural Network) types such as RBF (Radial Basis Function networks) and MLP (Multi-Layer Perceptrons) for predicting yields from sunflower plant characteristics.

MATERIALS

The city of Bingöl is located in the High Euphrates District over from 38° 27′ E to 40° 27′ E longitude and from 41° 20′ N to 39° 54′ N latitude. Generally, two types of soil are observed in Bingöl. Brown and brown-red soil is mainly observed in the sloping-rough lowlands, while alluvial soil is observed in the valleys; these soils can vary in terms of organic materials. The city generally has fertile soil. The area in which the trial was conducted has a coarse soil that has not been cultivated for years; the sample taken from the soil at a depth of between 0 to 30 cm was determined to be loamy (middle structure), with a pH of 6.37 (slightly acidic), having a salt content of 0.0315 % (saltless), with 1.905 % organic material (low) and 0.36 % lime (low lime). In addition, its beneficial amount of P205 (phosphates) was determined to be 7.91 kg/da (enough), and its K20 (potash) was determined to be 24.51 kg/da (low) (Anonymous, 2014).

Bingöl, and its associated region located in the eastern transition zone, has hot summers and cold winters because of its elevation and exposure to moist–cool air masses coming from the North.

In the research, ten varieties of sunflowers (Çiğdem 1, Sirena, Sanbro, Dkf 2525, Transol, Tr-3080, Vinimik, Confeta, Ege 2001 and Alhaja) supplied from various sources were used to determine the circumstances in the Eastern Anatolian region.

The research was based on an experimental design using triple randomized blocks. Each parcel was 4 meters in length and 2.8 in width, while the parcel area was 11.2 m^2. While each parcel consisted of 4 lines, inter-row distance was 70 cm, and the intra-row distance across was 25 cm. The whole experiment consisted of 30 parcels. The parcel intra-row distance was 25 cm and 2 seeds were planted every 25 cm. Thinning was carried out 3 or 4 weeks after commencing growth.

To prevent bird harm, perforated bags were raised over platforms in the middle two lines of all the land parcels. Bags were perforated with small holes to prevent platforms from going moldy and allowing the platforms to breathe.

During harvest time, the two lines in the middle were harvested while each line on the sides and the 0.4 m section on the edges were left as edge effects. Then, the harvested platforms were desiccated and sunflower grains were removed by hand.

The measured characteristics of the sunflowers in the study were yield per decare (YPD), plant length (PL), number of nodes (NN), stalk circumference (SC), table diameter (TD), weight of 1000 seeds (SW1000), kernel percentage (KP), oil production rate (OR) and protein ratio (PR). Descriptive statistics for the dependent and independent variables are given in Table 1.

Table 1. Description of the dependent and independent variables (means and standard deviations).

	N	$\bar{X}$	S
PL (cm)	30	138.923	12.456
NN (number/plant)	30	23.853	2.259
SC (mm)	30	19.933	1.431
TD (cm)	30	18.277	0.850
SW 1000 (g)	30	61.410	15.079
KP (%)	30	67.103	7.923
YPD (kg/decare)	30	237.055	36.224
OR (%)	30	36.089	5.054
PR (%)	30	31.212	2.302

$\bar{X}$: Mean, s: Standard Deviation

METHODS

CHAID is an analysis based on a criterion variable with two or more categories (Chen, 2003; Díaz-Pérez *et al*., 2005; Legohérel *et al*., 2015), and it is a non-parametric method. Nominal type and interval variables can be considered as predictors in CHAID. Continuous variables can be chosen as criterion variables (Diepen and Franses, 2006).

Exhaustive CHAID was considered by Biggs *et al*. (1991). Biggs suggested finding the best split by merging similar pairs continuously until only a single pair remained. The set of categories with the largest significance is the best split for that predicted variable.

CART provides graphic representations in which a discriminating criterion is used to split a sample into sub-groups of elements with shared characteristics that are depicted as nodes and branches (Pérez 2006). Minimum node size was determined to be 4:2 in this study. The minimal node process continues until each node reaches a user-specified minimum node size and becomes a terminal node. The computational complexity of the recursive partitioning algorithm used for growing regression trees is highly dependent on the choice of the best split for a given node. Ten-fold cross-validation and the one standard error rule were used to find the best tree. The inherent stopping criteria are provided by the tree-building algorithm itself, where it stops splitting the node when all the samples in the node give the same response.

A multilayer perceptron (MLP) network has one or more hidden layers of neurons followed by an output layer of linear neurons. The back-propagation algorithm uses supervised learning where a set of inputs and outputs is provided for the network. The linear output layer lets the network produce values outside the range −1 to +1. The conversion of the network outputs into the proper values of the output variable takes place during post-processing. The post-synaptic potential function of the artificial neurons in ANN that implements the back-propagation algorithm is a weighted sum of the inputs x_i and their respective weights w_{ij} (see equation 1).

$$A_j(\bar{x},\bar{w}) = \sum_{i=0}^{n}(x_i, w_{ij}) \quad (1)$$

The value of this function becomes, in turn, an argument of the activation function.

The most common output function of back-propagation is the sigmoid function as shown in equation 2 (Pal *et al*., 2015).

$$O_j(\bar{x},\bar{w}) = \frac{1}{1+e^{-A_j(\bar{x},\bar{w})}} \quad (2)$$

The Radial Basis Function (RBF) network is a forward Neural Network category composed of three layers: an input layer, a hidden layer, and an output layer (Yu and He, 2006). In RBF networks, outputs are determined by calculating the distance between network inputs and the centers of the hidden layer. The second layer is the hidden linear layer, and outputs of this layer are weighted forms of the outputs of the input layer. Each hidden layer neuron with a vector parameter is called the center. Therefore, a general description of the network is given by equation (3) (Robert and Howlett, 2001).

$$\hat{y} = \sum_{i=1}^{I} w_i \phi(\|x - c_i\|) + \beta \quad (3)$$

The standard mode is usually the Euclidean distance, and RBF implements Gaussian functions as shown in equation (4).

$$\varphi(r) = exp(\ \alpha_i \|x\ \ c_i\|^2) \quad (4)$$

In equations (3) and (4), the following definitions are considered: $i \in \{1, 2, 3, \ I\}$, so I is the number of neurons in the hidden layer; w_i, the weight between a neuron in the hidden layer and the output; φ, is the Gaussian function; *ai*, the spread parameter (amount of variance) neuron; *x*, the input data vector; c_i, the center vector of the neuron; β, bias of the output; $\hat{y}$, output of the network.

The construction, training and testing of ANNs were conducted to enable the choice of the optimal network structure and the number of neurons in the hidden layers. The final criterion was applied to find the best network, like the lowest RMSE on the validation set.

The log-sigmoid transfer function was used for nodes in the hidden layer of these networks, while two different transfer functions were utilized in the output layer (Herzog, 2006). Radial centers and deviations of the RBF network were determined by Gaussian activation functions (*h(x)*). Parameters of *h(x)* are *r* (the radius or standard deviation) and *c* (the center or average taken from the input space) defined separately for each RBF unit. The function *h (x)* is as follow:

$$h(x) = exp\left(-\frac{(x-c)^2}{r^2}\right)$$

The epoch number was set at 1000 for the MLP algorithms. Sixty different configurations of ANN were investigated. Perturb and profile methods were used to determine the influence of each input variable and its contribution to the output of the ANN model.

The plant characteristic measures were exposed to MLP and RBF ANN types for dependent variable prediction in a training–testing set proportion of 80:20.

Statistical evaluations for ANN were carried out using IBM SPSS 23 software.

The formulas for statistical error analysis to compare the predictive performance of the algorithms are shown in equations 5 to 8 (Grzesiak and Zaborski, 2012; Ali, *et al*., 2015). Definitions of RRMSE are based on Gandomi and Roke (2013) and shown in equation 9.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \quad (5)$$

$$Adj.R^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \quad (6)$$

Global relative approximation error (RAE),

$$RAE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} Y_i^2}} \quad (7)$$

Standard Deviation Ratio,

$$SD_{ratio} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (8)$$

Relative root mean squared error (RRMSE),

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2} \times 100 \quad (9)$$

RESULTS

Yield per decare: To determine the factors affecting yield per decare in the sunflower plant, various data mining algorithms and artificial neural net methods were compared in accordance with different goodness of fit criterion (Table 2).

Table 2. Predictive performance of CART, CHAID, Exhaustive CHAID and ANN types (yield per decare).

	CHAID	Exhaustive CHAID	CART	MLP	RBF
RRMSE	0.00036	0.00025	0.00036	0.00053	0.00032
SD ratio	0.582	0.403	0.582	0.844	0.523
RAE	0.086	0.060	0.086	0.125	0.078
R^2	0.661	0.837	0.745	0.453	0.726
Adjusted R^2	0.649	0.825	0.736	0.433	0.706
r	0.813	0.915	0.868	0.673	0.852

As seen in Table 2, the Exhaustive CHAID algorithm, which has the biggest r, R^2, Adj. R^2 values, and the lowest RRMSE, RAE and SD ratio values, was seen to be the most suitable algorithm. The regression tree diagram made in accordance with the Exhaustive CHAID algorithm is shown in Figure 1.

The minimum *parent node*:*child node* ratio was determined to be 4:2. The node on the top of the regression tree is node 0, which means the root note, and consists of all the plants (n=30). The average yield per decare in the sunflowers was determined to be 237.055 kg, while the standard deviation was calculated to be 36.224 kg. Node 0, the root node, was divided into 6 new child nodes (Node 1, 2, 3, 4, 5 and 6) according to the oil production rate.

In the regression tree, it is observed that as we proceed from node 2 to 4, and as the oil production rate (28.040 to 38.860) increases, the yield per decare increases as well (190.973 to 248.130 kg). This case does not mean that the sunflower that has highest oil production rate (%) always has the highest yield per decare. The average sunflower yield per decare (standard deviation) values are respectively determined to be 290.155 (S=8.971) kg for node 1, 190.973 (S=7.289) kg for node 2, 224.500 (S=26.561) kg for node 3, 248.130 (S=7.577) kg for node 4, 211.987 (S=19.929) kg for node 5 and 253.660 (S=7.991) kg for node 6.

Among the variables in the model, the oil production rate (%) (Adj. P=0000, F=16.994), weight of 1000 seeds (g) (Adj. P=0.011, F=20.417) and kernel percentage (%) (Adj. P=0.005, F=54.226) are determined to influence the yield per decare.

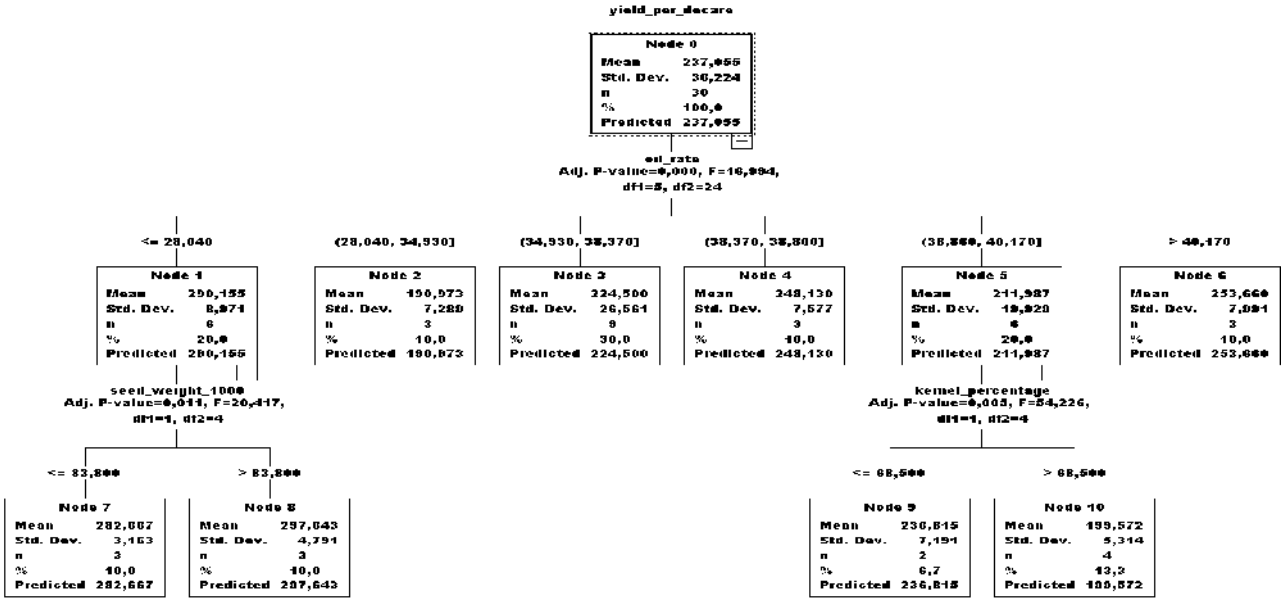

Figure 1. Regression tree diagram for yield per decare using the Exhaustive CHAID algorithm

The plant characteristics data was assessed by MLP and RBF ANN types for the yield per decare estimation at a training–testing set proportion of 80:20. The order of importance for the independent variables for the MLP algorithm was: protein ratio (100%) > kernel percentage (73%) > stalk circumference (49.6%) > plant length (36%) > weight of 1000 seeds (32.7%) > number node (32.2%) > oil production rate (10.5%) > table diameter (6.5%). The order of importance for the RBF algorithm was: weight of 1000 seeds (100%) > kernel percentage (88.7%) > plant length (61.5%) > oil production rate (61.5%) > stalk circumference (58.8%) > number node (53.1%) > table diameter (34.3%) > protein ratio (18.3%).

Oil Production Rate (%): The minimum parent node: child node ratio was determined to be 4:2. The model quality values calculated according to the regression tree are given in Table 3.

When the decision tree diagram made in accordance with the CART algorithm was examined, it was determined that the first effective independent variable on the fat rate was the grain ratio, the second was the kernel percentage and plant length, the third was the handle periphery and grain ratio, the fourth was the 1000 seed weight and the fifth was yield per decare.

The plant characteristics data was assessed by multilayer perceptron (MLP) and radial basis function (RBF) ANN types for oil production ratio (%) estimation at a training–testing set proportion of 80:20. The order of importance for independent variables for the MLP algorithm was weight of 1000 seeds (100%), kernel percentage (48.8%), number node (17.4%), protein ratio (16.4%), plant length (8%), table diameter (5.3%), yield per decare (4.8%) and stalk circumference (3.9%), respectively. The order of significance for the RBF algorithm was kernel percentage (100%), weight of 1000 seeds (97.5%), protein ratio (67.7%), yield per decare (63%), number node (11.9%), stalk circumference (8.4%), table diameter (7.6%) and plant length (4%), respectively.

Table 3. Predictive performance of CART, CHAID, Exhaustive CHAID and ANN types (for Oil production rate).

	CHAID	Exhaustive CHAID	CART	MLP	RBF
RRMSE	0.00081	0.00061	0.00045	0.00114	0.00153
SD ratio	0.216	0.162	0.155	0.303	0.414
RAE	0.029	0.022	0.021	0.041	0.056
R^2	0.953	0.974	0.976	0.908	0.832
Adjusted R^2	0.951	0.973	0.975	0.905	0.826
r	0.976	0.987	0.988	0.953	0.912

Plant length: The minimum parent node: child node ratio was determined to be 4:2. The model quality values calculated according to the regression tree are given in Table 4.

When the regression tree diagram constructed according to the exhaustive CHAID algorithm was examined (Figure 3), it was determined that only the grain ratio affected the plant length (Adj. P=0.000, F=30.337).

The node at the top of the regression tree is node 0 (the root node) consisting of all the observations (n = 30). It was estimated that the average sunflower plant length was 138.923 cm and the standard deviation was 12.456 kg. Seven child nodes (node 1 to node 7) were allocated for the node 0 kernel percentage (%).

The order of importance of independent variables for the MLP algorithm was number node (100%), protein ratio (38.2%), head width (23.7%), kernel percentage (18.9%), weight of 1000 seeds (12.3%), stalk circumference (9.3%), yield per decare (9%) and oil production rate (2.6%), respectively. The order of significance for the RBF algorithm was number node (100%), head width (60%), stalk circumference (46.8%), oil production rate (36.8%), protein ratio (25.7%), yield per decare (16.8%), weight of 1000 seeds (4.6%) and kernel percentage (3.2%), respectively.

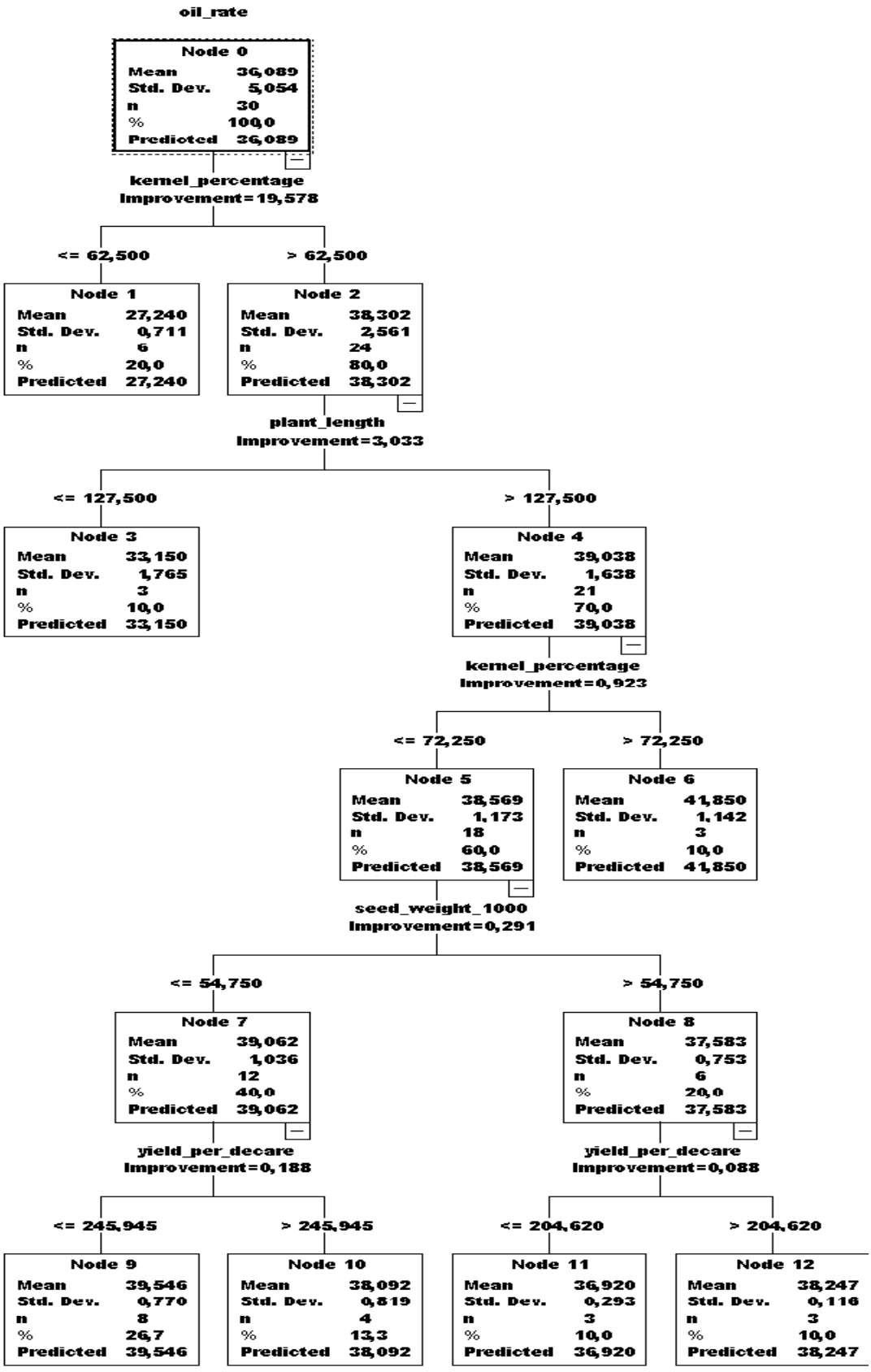

Figure 2. Decision tree diagram for oil production rate using the CART algorithm

Table 4. Predictive performance of CART, CHAID, Exhaustive CHAID and ANN types (plant length).

	CHAID	Exhaustive CHAID	CART	MLP	RBF
RRMSE	0.00129	0.00021	0.00101	0.00027	0.00025
SD ratio	0.409	0.335	0.362	0.425	0.391
RAE	0.036	0.029	0.032	0.037	0.034
R^2	0.832	0.887	0.869	0.826	0.846
Adjusted R^2	0.826	0.883	0.859	0.820	0.841
r	0.912	0.942	0.932	0.909	0.920

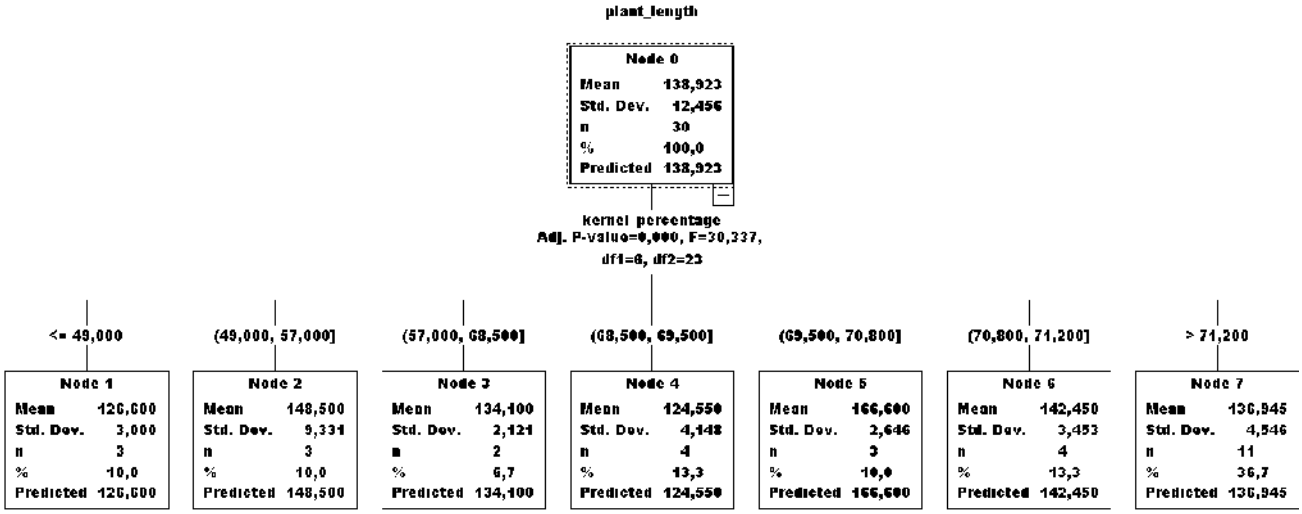

Figure 3. Regression tree diagram for plant length using the Exhaustive CHAID algorithm

DISCUSSION

In this study, SD ratio values for the applied algorithms: Exhaustive CHAID and CART, were calculated to be 0.403, 0.335 and 0.155, respectively. It could be advocated that the algorithm whose SD ratio was less than 0.40, or between 0 and 0.10, was a good fit or a very good fit (Grzesiak and Zaborski, 2012).

Göksoy *et al*. (2002) found that the number of seeds per head had the greatest direct effect (+0.7269) on seed yield in sunflowers, followed by the 1000-seed weights (+0.3215) and head diameter (+0.1689) using path analysis. The percentages for the direct effects on seed yield for the number of seeds per head, 1000-seed weights and head diameter were 80.8%, 50.6% and 24.0%, respectively,. Plant height showed the highest indirect effect on seed yield through the number of seeds per head (+0.4507 with a correlation of 74.6 %).

A randomized blocks experimental design was established with four replications by Ergen and Sağlam, 2005. In that experiment, some of the traits related to yield were examined, such as the plant height, the head diameter, the weight of 1000 seeds, the seed yield, seed length, the hull ratio, and the ratio of oil and protein. According to path analysis, it was determined that the direct influences of seed and plant height are especially significant for the seed yield and the protein ratio in edible sunflowers

Coşge (2007), determined the effect of foliar applied methanol on some morphological traits and the seed yield of sunflowers (*Helianthus annuus L*.) in an experiment established with a completely randomized design in split plot arrangements with three replications. Varieties were affected only in plant height at harvest time, and in leaf numbers and lengths at the beginning of flowering.

Differences were found in seed length, head diameter, plant height at the beginning of flowering, the thousand seed weights and the seed yield/ C 207 plants for different treatments. Methanol treatments positively affected the head diameter and thousand seed weights.

Kaya *et al*. (2009), found significant correlations between fat yield and other items in regression analysis studies to determine the relationship between yield factors playing an important role in the formation of the fat yield in sunflowers; the highest coefficient was grain yield, and this was followed by plant length, the weight of 1000 seeds and head diameter. In the formation of fat yield, the grain yield was more dominant than the fat ratio, the linearity in the grain yield and the plant length; quadratic relationships were determined in others. A quadratic relationship was determined in relation to the head diameter and the weight of a 1000 seed size.

According to the results of the regression analyses, it was found that the fat yield would be compromised when more than 70 grams of grain, more than 53% of oil and 24 cm of table diameter were required.

Mnayer *et al*. (2017) obtained a second-order polynomial equation for estimating absolute yield (expressed in g per 100 g of dry weight) using applied ultrasonic power, sonication time and temperature in coded units. In summary, the results of this study and earlier studies could not be applied due to the use of different plant characteristics, sample size and statistical analysis techniques. The results of other methods applied to this plant have been shown in this study.

Conclusion: The results obtained for the methods mentioned are summarized as follows:

1. On average, the highest yield (297.643 kg) was derived from the subgroup of plants with oil production rate ≤ 28.400 and weight of 1000 seeds > 83.800.
2. The highest oil content (41.850) was obtained with a kernel percentage > 62.500, a plant length > 127.5, and a kernel percentage > 72.250.
3. The longest plant length (166.600 cm) was produced when 69.500 < kernel percentage ≤ 70.800.

In conclusion, the decare yield and plant length of the sunflower plant were best predicted by the Exhaustive CHAID algorithm and the oil percentage (%) by the CART algorithm.

REFERENCES

Ali, M., E. Eyduran, M. M. Tariq, C. Tirink, F. Abbas, M. A. Bajwa, M. H. Baloch, A. H. Nizamani, A. Waheed, M. A. Awan, S. H. Shah, Z. Ahmad, and S. Jan (2015). Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. Pakistan J. Zool., 47(6): 1579-1585.

Anonymous (2014). Turkish Statistical Institute.

Arıoğlu, H. (1999). Yağ Bitkileri Yetiştirme ve Islahı. Ç. Ü. Ziraat Fakültesi Genel Yayın No: 220, Adana.

Bertoria L, R. Burak, and A. Nivio (1998). Effect of plant densities on yield and quality of for age maize. Maize growers co-operative news letter santa catarina, Brazil.

Biggs, D., B. De Ville, and E. Suen (1991). A method of choosing multiway partitions for classification and decision trees. J. Applied Statistics, 18(1): 49–62.

Chen, J. S. (2003). Market segmentation by tourists' sentiments. Annals of Tourism Research, 30(1): 178–193.

Coşge, B. (2007). Effect of Methanol on Some Morphological Characters and Seed Yield of Sunflower (*Helianthus annuus L.*). Tarım Bilimleri Dergisi, 13(3): 246-252.

Díaz-Pérez, F. M., M. Bethencourt-Cejas, and J. A. Álvarez-González (2005). The segmentation of Canary island tourism markets by expenditure: Implication for tourism policy. Tourism Management, 26(6): 961–964.

Diepen, M.van, and P. H. Franses, (2006). Evaluating chi-squared automatic interaction detection. Information Systems, 31: 814–831.

Ergen, Y., and C. Sağlam (2005). Yield And Yield Characters of Different Confectionery Sunflower Varieties in Conditions of Tekirdag. J. Tekirdag Agricultural Faculty, 2(3): 221-227.

Gandomi, A. H., and D. A. Roke (2013). Intelligent formulation of structural engineering systems. In: Seventh MIT Conference on Computational Fluid and Solid Mechanics- Focus: Multiphysics and Multiscale, 12-14 June, Cambridge, USA.

Göksoy, A. T., A. Türkeç, and Z. M. Turan (2002). Correlation and Path Analysis Between Seed Yield and Certain Yield Components in New-Improved Synthetic Varieties of Sunflower (*Helianthus annuus L.*). Uludağ Üniversitesi Ziraat Fakültesi Dergisi, 16: 139-150.

Göksu, Ç. (2007). Bitkisel Yağlar. T. C. Başbakanlık Dış Ticaret Müsteşarlığı İhracatı Geliştirme Etüt Merkezi.

Grzesiak, W., and D. Zaborski (2012). Examples of the use of data mining methods in animal breeding. (Book) ISBN 978-953-51-0720-0.

Herzog, M. A. (2006). Machine and Component Residual Life Estimation through the Application of Neural Networks. Department of Mechanical and Aeronautical Engineering University of Pretoria, South Africa, Master of Thesis, 150p.

Kaya, Y., G. Evci, V. Pekcan, T. Gücer, S. Durak, and A. Üstün (2005). Çerezlik Ayçiçeğinde Bazı Köy Çeşitleri ve Hibritlerinin Performanslarının Değerlendirilmesi. 6. Tarla Bitkileri Kongresi, Antalya, Türkiye, 2: 619-622.

Kaya, Y., G. Evci, V. Pekcan, T. Gücer, and M. İ. Yılmaz (2009). The Determination Relationships Between Oil Yield and Some Yield Traits in Sunflower. Tarım Bilimleri Dergisi, 15(4): 310-318.

Legohérel, P., C. H. C. Hsu, and B. Daucé (2015). Variety-seeking: Using the CHAID segmentation approach in analyzing the international traveler market. Tourism Management, 46: 359–366.

Mnayer, D., A. S. Fabiano-Tixier, E. Petitcolas, K. Ruiz, T. Hamieh, and F. Chemat (2017). Extraction of green absolute from thyme using ultrasound and sunflower oil. Resource-Efficient Technologies 3: 12–21.

Mojiri, A. and A. Arzani (2003). Effects of nitrogen rate and plant density on yield and yield components

of sunflower J. Sci and Technology of Agric. And natural resources, 7(2): 115-125.

Pal, A., J. P. Singh, and P. Dutta (2015). Path length prediction in MANET under AODV routing: Comparative analysis of ARIMA and MLP model. Egyptian Informatics J., 16: 103-111.

Pérez, J. M. (2006). Árboles consolidados: Construcción de un árbol de clasificación basado en múltiples submuestras sin renunciar a la explicación Unpublished PhD. thesis. Universidad del País Vasco.

Poyraz, O. (2012). Farklı Olgunlaşma Grubundaki Hibrit Ayçiçeği (*Helianthus annuus L.*) Çeşitlerinin Verim ve Kaliteleri Üzerine Bitki Sıklığının Etkisi. Yüksek Lisans Tezi, Namık Kemal Üniversitesi, Tekirdağ.

Robert J. and L.C.J. Howlett (2001). Radial basis function networks 2: New Advances in Design. Physica-Verlag; Herdelberg, Germany: ISBN: 3790813680.

Vasudevan P., S. Kashyap and S. Sharma (1997). Tagetes: a multipurpose plant. Bioresource Technology, 62(1-2):29–35.

Yu, B. and X. He (2006). Training radial basis function networks with differential evolution. In: IEEE Conference on Granular Computing; p. 934-941.