Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 13(2), 1323-1334, 2023
Journal of the Institute of Science and Technology, 13(2), 1323-1334, 2023

ISSN: 2146-0574, eISSN: 2536-4618

Agricultural Economy **DOI: 10.21597/jist.1177194**

Research Article

Received: 19.09.2022 **Accepted: 11.11.2022**

Atıf İçin: Karadaş, K., Kadirhanoğulları, İ. H. ve Konu Kadirhanoğulları, M. (2023). Karpuz (Citrullus lanatus (Thunb.) Matsum. & Nakai) Üretimini Etkileyen Faktörlerin Veri Madenciliği ile Tahmini. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(2), 1323-1334.

To Cite: Karadaş, K., Kadirhanoğulları, İ. H. & Konu Kadirhanoğulları, M. (2023). Prediction of The Factors Affecting Watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) Yield Using Data Mining. *Journal of the Institute of Science and Technology*, 13(2), 1323-1334.

Karpuz (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) Üretimini Etkileyen Faktörlerin Veri Madenciliği ile Tahmini

Köksal KARADAŞ[1*], İbrahim Hakkı KADİRHANOĞULLARI[2], Meryem KONU KADİRHANOĞULLARI[3]

Öne Çıkanlar:
- Karpuz üretiminin tahmin edilmesi
- Karpuz verimini etkileyen föktörler
- Karpuz verimi ve veri madenciliği

Anahtar Kelimeler:
- Veri madenciliği
- Regresyon ağacı
- Karpuz
- Diyarbakır

ÖZET:

Bu çalışmanın amacı Diyarbakır ilinde karpuz verimini etkileyen faktörleri belirlemektir. Veri Madenciliği Regresyon Ağacı yönteminden Ki-kare otomatik etkileşim detektörü (CHAID) algoritması kullanılan çalışmanın verileri Basit Tesadüfi Örnekleme Yöntemi'ne göre belirlenen 80 adet karpuz üreticisinden elde edilmiştir. Oluşturulan model de Bağımlı değişken WY (karpuz verimi), bağımsız değişkenler R (bölge), AF (çiftçinin yaşı), EL (eğitim düzeyi), CA (ekim alanı), FD (gübreleme zamanı), FA (gübre miktarı), DS (ilaçlama zamanı), AS (ilaç miktarı), NI (sulama sayısı), IT (sulama süresi), AN (çapa sayısı) ve HT (hasat zamanı) olarak belirlenmiştir. Dekar başına ortalama 4488.9 kg karpuz verimi elde edilmiş ve çapa sayısı karpuz verimini en çok etkileyen değişken olmuştur. Sonuç olarak birim alandan daha yüksek verim alabilmek için karpuz üreticilerinin 4 defadan fazla çapa, 2 saatten az olmak üzere 5-6 defa sulama yapmaları ve Mayıs ayında gübre uygulamaları yapmaları gerekmektedir. Ayrıca Çermik, Eğil, Yenişehir ve Bismil karpuz üretimi için daha uygun bölgeler olarak belirlenmiştir.

Prediction of The Factors Affecting Watermelon (Citrullus lanatus (Thunb.) Matsum. & Nakai) Yield Using Data Mining

Highlights:
- Forecasting watermelon production
- Watermelon yield and data mining
- Factors affecting watermelon yield

Keywords:
- Data mining
- Regression tree
- Watermelon
- Diyarbakır

ABSTRACT:

The aim of this study was to evaluate the factors of affecting watermelon yield in Diyarbakır province. The data was obtained from surveying of 80 watermelon farmers in Diyarbakır province, Turkey by Simple Random Sampling Method using the Chi-square automatic interaction detector (EXHAUSTIVE CHAID) algorithm of the Data Mining Regression Tree methods. In the model created, the dependent variable was WY (watermelon yield), and the independent variables were determined as R (region), AF (age of farmer), EL (education level), CA (cultivation are), FD (fertilization date), FA (amount of fertilization), DS (date of spraying), AS (amount of spraying), NI (number of irrigation), IT (irrigation time), AN (anchor number), HT (harvest time). As a result of the study, the factors that significantly affect the yield of watermelon; AN, NI, HT, CA, R has been determined. An average of 4488.9 kg watermelon yield per decare was obtained and the number of hoes was the variable that most affected the watermelon yield. As a result in order to get a higher yield per unit area, watermelon producers should anchor number more than 4 times, irrigate 5 to 6 times at less than 2 hours, and apply fertilizer in May. In addition, Çermik, Eğil, Yenişehir and Bismil were determined as more suitable regions for watermelon production.

[1*] Köksal KARADAŞ (**Orcid ID:** 0000-0003-1176-3313), Iğdır Üniversitesi, Ziraat Fakültesi, Tarım Ekonomisi Bölümü, Iğdır, Türkiye

[2] İbrahim Hakkı KADİRHANOĞULLARI (**Orcid ID**: 0000-0002-9640-8910), Iğdır Üniversitesi, Uygulamalı Bilimler Fakültesi, Organik Tarım İşletmeciliği Bölümü, Iğdır, Türkiye

[3] Meryem KONU KADİRHANOĞULLARI (**Orcid ID:** 0000-0001-7359-7061), Kafkas Üniversitesi, Sosyal Bilimler MYO, Çocuk Gelişimi Bölümü, Kars, Türkiye

***Corresponding Author:** Köksal KARADAŞ, e-mail: kkaradas2002@gmail.com

Ethics Committee Approval: Bu makalede yer alan anket çalışması için "Iğdır Üniversitesi Bilimsel Araştırma ve Yayın Etik Kurulu Başkanlığı'nın Tarih: 25.01.2021 Toplantı No 2021-04 sayılı kararı ile Etik Kurul Onayı almıştır.

INTRODUCTION

Vegetables are very important for human nutrition due to their abundant mineral, vitamin and antioxidant contents (Özçınar, 2020). The origin of watermelon, of which cultivation dates back for a long time, is Africa (Wehner, 2010; Erincik, 2015). Watermelon *(Citrullus lanatus* (Thunb.) Matsum. & Nakai) and melon (*Cucumis melo* L.), which are the members of *Cucurbitaceae* family and are highly economically important species, are cultivated in warm regions (Güner and Wehner 2004; Öztürk, 2018).

The lycopene content, that prevents the occurrence of some diseases, of watermelon is higher compared to the other red-colored fruit species (Güner and Wehner 2004; Öztürk, 2018). Watermelon contains 8-14% sugar, while it is poor in protein, fat, and calories. However, in addition, the watermelon is rich in vitamins B, C and A, and contains calcium (Ca), iron (Fe), phosphorus (P) and magnesium (Mg) (Anonymous, 2005; Öztürk, 2018). Watermelon, which consists of 95% water, is usually consumed fresh, but also used in the production of syrup, fruit juice, jam, pickle or canned food. The seeds of watermelon are used in the cosmetic and pharmaceutical industry, and the shells can also be used in animal nutrition (Güneş and Aşkın, 2016). Watermelon is named as New world watermelon, Mottled watermelon, Black watermelon (black ice watermelon), Gülle watermelon, Washington watermelon and Diyarbakır watermelon based on the shape, color and, place of cultivation (Altaş, 2015).

Watermelon production of the world in 2021 was 104 136 387 million tons, and China ranked the first with 63 024 614 million tons and followed by Turkey with 4 031 174 million tons of production (Anonymous, 2021). Watermelon and melon are among the most produced species of the cucurbit family, which has a share of 40% in vegetable production in Turkey (Çat et al., 2016; Özçınar, 2020). Watermelon cultivation is carried out in 73 provinces of Turkey, and the highest production in 2021 was carried out in Adana province with 689 212 tons, while Diyarbakır ranked the 7th with 121 749 tons of watermelon production. Antalya, İzmir, Mardin and Şanlıurfa provinces are the other provinces where watermelon is produced the most in Turkey (TUİK, 2021).

Watermelon cultivation is mostly carried out in greenhouses in Adana, Antalya and Mersin, so the amount of watermelon production in open field conditions in Diyarbakır is too important to be underestimated. Watermelon has a special importance for Diyarbakır province, where watermelon has been produced for a long time and, it is identified with the watermelon (Koçkaya, 2019). Diyarbakır watermelon is grown in loamy, sandy and, pebbly lands next to the Tigris River with methods unique to Diyarbakır and has an important place in social, folkloric, and economic life of the region. The unique feature of Diyarbakır watermelon is its size and taste (Okumuş, 2010).

Studies on watermelon yield reported that main vine length and number of side branches affect watermelon yield (Ban et al., 2011), and watermelon yield can be increased at different spacing and mulching levels (Sylvestre et al., 2014). Similarly, the decrease in plant density increased fruit weight and fruit yield of watermelons (Kavut et al., 2014), and watermelon yield in irrigated fields (37.28 tons/ha) was higher than the yield (9.98 tons/ha) recorded in non-irrigated lands (Pejic et al., 2016). According to in addition, Filho et al. (2019) stated that application of 120 kg/ha NPK fertilizer increased watermelon yield, nutrient accumulation and fruit quality. In another study, watermelon yield significantly increased with direct application of nitrogen in drip irrigation system (Rolbiecki et al., 2020).

Data Mining Decision Tree models are used recently in evaluating agricultural production and yield data, determining, and evaluating product, input, and yield relationships. Decision Tree Model has been mostly used in livestock farming (Aytekin et al., 2018; Celik et al., 2018; Karadas and Birinci, 2019; Zaborski et al., 2019), while studies on crop production are very limited (Küçükönder et al., 2015;

Irmak and Ercan, 2017; Bostancı and Eren Atay, 2018). The aim of this study is to determine the factors affecting watermelon yield in Diyarbakır province by using Data Mining Regression Tree Method.

MATERIALS AND METHODS

Data Collection Phase

The material of the research consists of the data obtained from the questionnaires made with 80 watermelon producers registered in the Agricultural Information System (AIS) of Diyarbakır Provincial Directorate of Agriculture and Forestry, Department of Crop Production. In addition, other studies related to the subject, domestic and foreign publications, records and statistical data of public institutions and organizations were also used in the study. The surveys were conducted between August and September 2019.

Diyarbakir province, which ranks 7th in Turkey's watermelon production (TUİK, 2021), have been selected as the study area. (Fig.1). Diyarbakır province is in located the central part of Southeastern Anatolia Region of Turkey, lies between 37° 56' latitude and 40° 13' longitude and the altitude is average 650 m (Anonymous, 2019a)

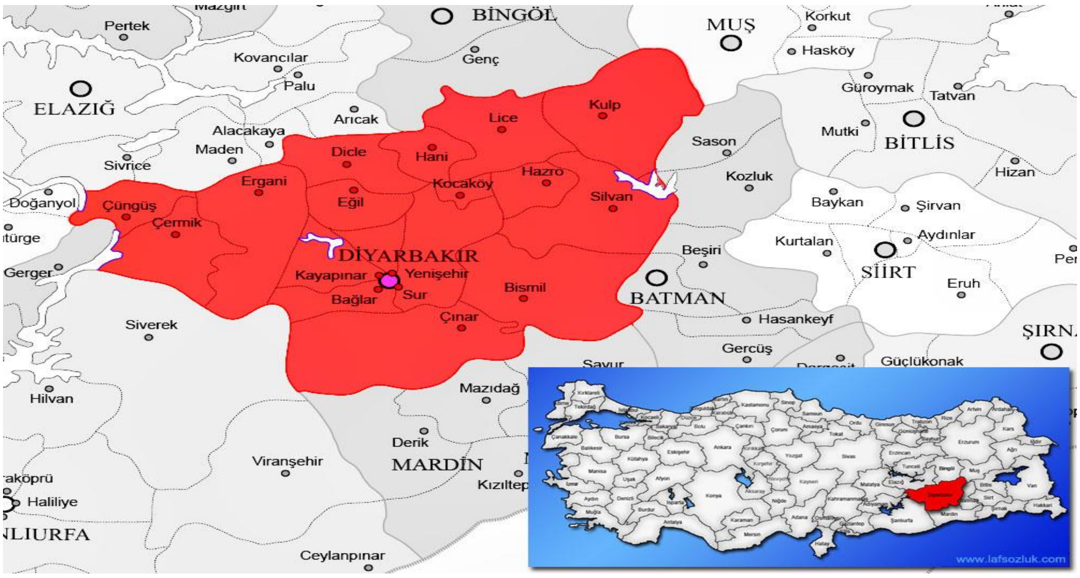

Figure 1. Location of study area

There were 326 watermelon growers registered in the Agricultural Information System (AIS) in Diyarbakır province (Anonymous, 2019b), and 80 producers were selected with "Simple Random Sampling Method" using the Equation 1 (Yamane, 2010) to best represent the watermelon grower's population. The data obtained in the surveys conducted with the producers were evaluated. In the evaluation, the watermelon production areas of the growers were taken into consideration.

$$n = \frac{NS^2}{(N-1)D^2 + S^2} \quad (1)$$

Where, n is the number of persons to be sampled, N is the size of surveyed farmers (326 producers), S^2 is the variance of the persons (90.06) and D is the correction factor, that has been calculated using following equation (Eq. 2).

$$(D) = (E/t)^2 \quad (2)$$

In the equation, coefficient of t is accepted as 1.6445 within 90% of confidence interval. E is the error term (1.525), and it is 10% of the average of size population.

$$n=\left(\frac{326\times(9.49)^2}{(326-1)\times(1.525/1.6445)^2+(9.49)^2}\right)=80$$

The sample size chosen was considered sufficient and the questionnaire was applied to 80 people. The districts where the surveys were covered and the number of the surveys are given in Table 1.

Table 1. The number of farmers and questionnaires by districts

	Number of Farmers	%	Number of Questionnaires
Sur	7	2	2
Çermik	93	29	23
Çınar	13	4	3
Dicle	9	3	2
Eğil	52	16	13
Ergani	73	22	18
Kayapınar	48	15	12
Kulp	5	2	1
Lice	11	3	3
Yenişehir	15	5	4
Total	326	100	80

Data Analysis Phase

Data mining was benefited as a method, and regression tree algorithm was used in the modeling phase of the study. Regression trees are advantageous algorithms for decision makers and are a data mining approach that is widely used for classification and prediction due to easy interpretation and comprehensibility (Karadas and Kadirhanogullari, 2017; Karakaya et al., 2018). CHAID (Chi-square automatic interaction detection), Exhaustive CHAID (exhaustive chi-square automatic interaction detection) and CART (Classification and Regression Tree) algorithms are effectively used in creating models with variables that have ordinal, nominal, and scale properties. The CART algorithm allows the creation of a decision tree structure based on binary split criteria by repeatedly splitting a node into two sub-nodes (Duru et al., 2017; Eyduran et al., 2017). The higher similarity rate among the members of result groups indicates that the splitting is more successful (Sun and Hui, 2008).

Some goodness of fit criteria reported by Grzesiak and Zaborski (2012) and Koç et al. (2017) were used to compare the regression tree algorithm performances (Table 2).

Where,

WY=Watermelon yield (kg/da)

R=Region (1:Çermik, 2:Sur, 3:Eğil, 4:Kayapınar, 5=Yenişehir, 6:Bismil)

AF=Age of farmer

EL=Education level, (1:illiterate, 2:literate, 3:primary school, 4:secondary school 5:high school 6:two-year degree, 7:bachelor's degree)

CA=Cultivation area (decare)

FD=Fertilization date (1:January, 2:February, 3:March, 4:April, 5:May)

FA=Amount of fertilization (kg/da)

DS=Date of spraying (1:January, 2:February, 3:March, 4:April, 5:May)

AS=Amount of Spraying (ml/da)

NI= Number of irrigation

IT= Irrigation time (hour)

AN=Anchor number

HT=Harvest time (6:June, 7:July)

Table 2. Goodness of Fit Criteria

Coefficient of Determination	$R^2 = \left[1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right] * 100$
Adjusted Coefficient of Determination	$\bar{R}^2 = \left[1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right] * 100$
Coefficient of Variation	$CV = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\bar{Y}}} * 100$
Standard Deviation Ratio	$SD = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$
Relative Approximation Error	$\text{RAE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} Y_i^2}}$
Root-mean-square error	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$
Mean absolute percentage error	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{Y_i - \hat{Y}_i}{Y_i}\right\rvert * 100$

$\varepsilon_i = Y_i - \hat{Y}_i$ (Actual WYF – Predicted WYF); WYF is watermelon yield factors, n is the number of watermelon (n), i………………. 1,2,…, n (treatment), $\bar{Y}$ ……………… overall mean, $\hat{Y}_i$……………..i. predicted value of watermelon, ε_i……………… chance related error value of watermelon, $\bar{\varepsilon}$……………… mean error value depending on chance error values, Y_i……………… I represent the WFY value of watermelon.

The goodness of fit criteria specified in Table 3 were used in the selection of the most suitable model. The criteria for goodness of fit were calculated in the R STUDIO program using the estimated values and observed (real) values of the dependent variable. Model comparisons were carried out based on the lowest RMSE, MAPE, RAE, coeffvar, SDratio and MAD values and the highest Pearson correlation coefficient, Rsq and ARsq values (Tatlıyer, 2020).

Table 3. Goodness of Fit of Regression Tree Algorithms for Factors Affecting Watermelon Production

	CHAID	EXHAUSTIVE CHAID	CART
Rsq	0.745	0.785	0.775
ARsq	0.670	0.722	0.709
R	0.863	0.886	0.880
SDratio	0.505	0.464	0.475
RAE	0.021	0.018	0.019
CV	15.240	13.990	14.330
MAPE	13.684	12.195	13.624
MAD	491.540	474.512	485.034
RMSE	679.902	624.129	639.193

The parent-child node was set to 6:3 and the cross-validation value to 10 to determine the highest predictive performance.

RESULTS AND DISCUSSION

While evaluating the results, first of all, demographic data of watermelon producers were evaluated. The most crowded age group of watermelon farmers was between 15 and 49 (71.3%) with 57 people, and the literate group was the major crowded group in education status with 53 people (66.3). The highest number of individuals in the family was between 3 and 7 with 36 farmers (45.0%), and the average density for the number of individuals in a family was 5 individuals. The number of producers who do not have an additional income was 44 (55%), other producers earn additional income by working in private sector or plant and animal production. Sixty-three producers (77.5%) were landowner and only 19 producers (23.8%) had an agricultural insurance. Büyükkalay (2019) stated in her study titled Watermelon Production and Marketing Structure in Antalya Province that 84% of producers are between the ages of 30-60, 80% are secondary school graduates and 50% have a family size of 4 people. According to Seçer et al. (2020) In his study titled The Factors Affecting the Producers' Decision to Grow Fruit Tree and Their Expectations for the Future in Ağrı Province, 47% of producers were between the ages of 40-50, 74% were secondary school graduates, and 59% had family size less than 6 people. According to Simsek et al. (2020) found the average age of producers to be 54 and their education period to be 8 years in their study titled Determination of Approaches of the Stone Fruit Producers to the Phytopathological Problems of Uşak Province.

Table 4. Demographic Data of Watermelon Producers

		N	%
Age Range	15-49	57	71.3
	50-64	22	27.5
	65+	1	1.3
Education Status	Illiterate	1	1.3
	Literate	53	66.3
	Elementary School	4	5.0
	Secondary School	20	25.0
	High School	2	2.5
Number of family members	3.00	1	1.3
	4.00	25	31.3
	5.00	36	45.0
	6.00	15	18.8
	7.00	3	3.8
District	Çermik	10	12.5
	Sur	9	11.3
	Eğil	22	27.5
	Kayapınar	26	32.5
	Yenişehir	2	2.5
	Bismil	11	13.8
Additional Income	Livestock	11	13.8
	Crop and Animal production	10	12.5
	Private Sector	12	15.0
	Government official	3	3.8
	No additional income	44	55.0
Land Entity	Sharecropper	2	2.5
	Owner	62	77.5
	Rent and Property	16	20.0
Agricultural Insurance Holders	Yes	19	23.8
	No	61	76.3

Model Fit Criteria evaluation revealed that the most suitable model is the regression decision tree diagram formed by Exhaustive CHAID algorithm (Figure 2). Celik et al. (2017) CART model is the most suitable, Karadas et al. (2017) found the CHAID model as the most suitable, and Karadas and

Birinci (2019) MARS model is the most suitable model. The most important factors affecting the watermelon yield in the regression decision tree diagram were determined Anchor number (AN), number of irrigation (NI), harvest time (HT), cultivation area (CA), region (R), and irrigation time (IT). Gungor and Balkaya (2015) reported that plant density had a positive effect on watermelon yield. Kuşcu et al. (2015) stated that irrigation water affects the yield of watermelon.

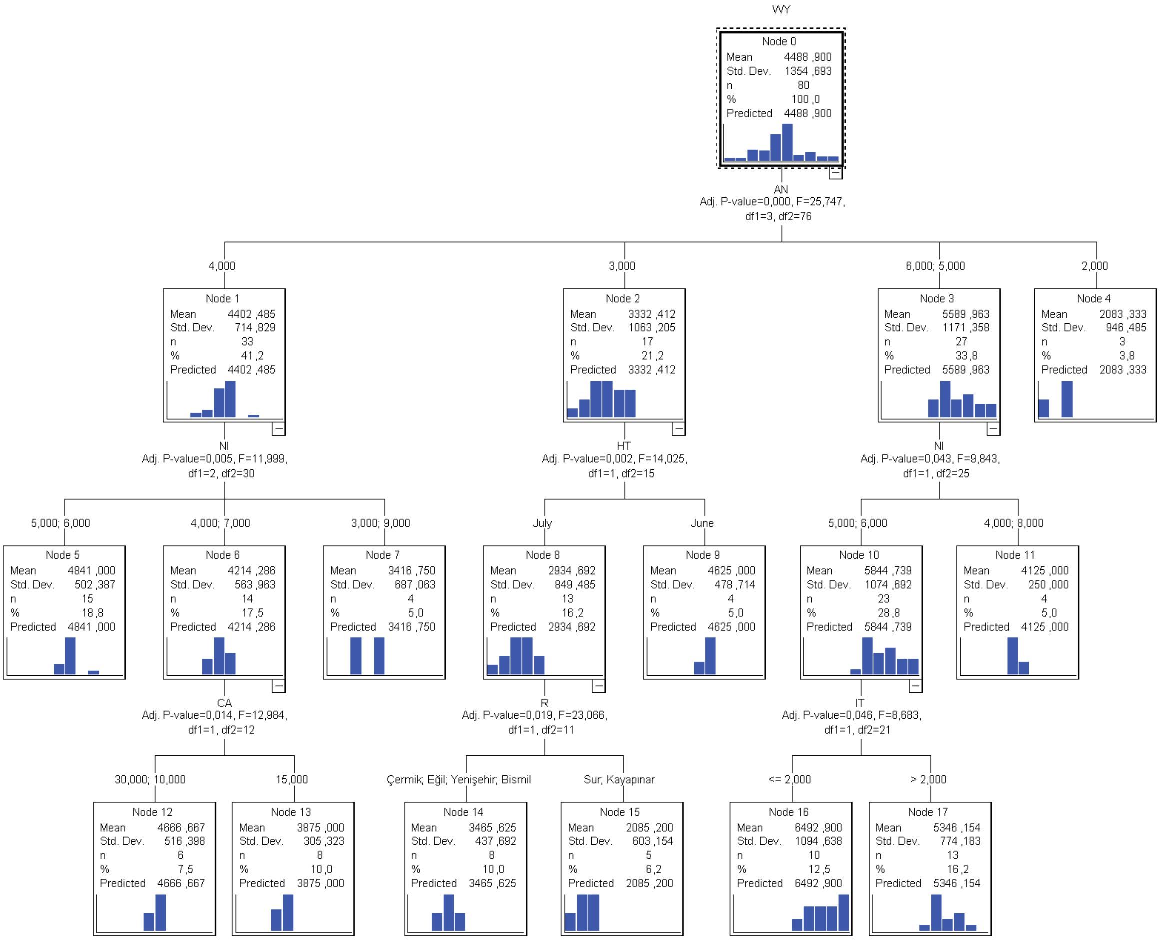

Figure 1. Regression Tree of Factors Affecting Watermelon Yield

The coefficient of determination estimated in the Exhausted CHAID algorithm was 79%. The regression decision tree diagram showed that an average watermelon yield (WY) of 80 producers who attended the questionnaire, in Node 0 was 4 488.9 kg/da (S:1 354). Watermelon yield determined in this study is higher than the WY (3 770 kg/da) determined by Tuna and Ozer (2005) in Muğla and the WY (3 832 kg/da) determined by Oktay and Doran (2005) in Diyarbakır province. The WY is like that (3 500-10 600 kg/da) reported by Koleboshina and Varivoda (2020) in the Volgograd region of Russia. However, the WY is lower than the WY (9 340 kg/da) determined by Yavuz et al (2020) in Konya Plain and the WY (19 400-24 900 kg/da) recorded by Malueva and Bocherova (2020) in Russia. Although studies have been conducted to determine WY under different regions and ecological conditions, data mining algorithms have not previously been used to determine the factors affecting WY. Therefore, it is estimated that this study will be one of the first studies in which data mining algorithms are used to determine the factors affecting WY.

The regression tree formed by the Exhaustive CHAID algorithm consists of three tree depths. Node 0 is divided into four child-nodes (Nodes 1, 2, 3 and 4) (Adj. P value=0.000, F=25.747, df1=3 and df2=76) according to the anchor number (AN). Node 1, where 33 watermelon producers applied 4 anchor (AN=4), constituted the first subgroup. Node 2 is the second subgroup (AN=3), and Node 3 comprise of the third group (AN=5 and 6). Node 4 is the terminal Node, which is the fourth subgroup, and is the Node 4 where the lowest WY was obtained. The average WY were 4 402 kg in Node 1 (S=714), 3 332 kg in Node 2 (S=1 063), 5 589 kg in Node 3 (S=1 171) and 2 083 kg in Node 4 (S=946), respectively. The hoeing in field crops is carried out for weed control, soil aeration and moisture conservation, and it increases the crop yield (Baydar, 2012). In addition, Şanlı et al. (2009) stated that hoeing and herbicide applications generally increased plant height in chickpea, where different herbicides and hoeing times were applied for weed control, compared to control plots. Node 3 in the decision tree indicated that the producers who applied 5-6 anchor obtained higher WY, therefore, 5-6 anchor operations in watermelon production can be recommended to obtain high WY.

Node 1 is divided into the fifth, sixth and seventh child-nodes as Nodes 5, 6 and 7 according to the NI. The Node 5 and 7 are the terminal nodes. The average WY of Node 5, 6 and 7 are 4 841 (S=502), 4 214 (S=563) and 3 416 (S=687) kg, respectively. Node 2 is divided into eighth and ninth child-nodes, based on the HT. Node 8 shows the producers harvesting watermelon in July and node 9 in June, and yields were determined as 2 934 (S=849) and 4 625 (S=478) kg, respectively. The differences in WY recorded in different time can be attributed to the effect of HT on WY. The producers harvesting in June obtain higher yield, therefore, harvesting in June can be recommended to the regional watermelon producers. The results of Tokgöz et al. (2015), who evaluated some quality parameters of two different watermelon cultivars grafted on three different rootstocks and harvested at two different times. It has been stated that the use of grafted seedlings provides an advantage in terms of mineral matter, and the samples taken from the second harvest period are more advantageous than the first harvest period.

Node 3 is divided into tenth and eleventh child-nodes as Nodes 10 and 11 based on the NI. The WY of Nodes 10 and 11 are 5 844 (S=1 074) and 4 125 (S=250) kg, and Node 11 is the terminal Node. Producers who irrigate 5 and 6 times obtained the highest yield. Producers irrigating less or more than 5-6 times obtained less yield. The insufficient water delays the maturation of watermelon plants, while excess water accelerates the vegetative growth and negatively affects the yield (Anonymous, 2020). In addition, Vural and Dağdelen (2008) stated the importance of irrigation intervals including water levels on the number of leaves per plant and on yield and agronomic characteristics of watermelon. The most appropriate irrigation is needed for higher yields, therefore 5-6 times irrigation can be recommended in watermelon cultivation to obtain optimum yield in the region.

Node 6 is divided into the twelfth and thirteenth child-nodes as Node 12 and Node 13 based on cultivation area, and the resulting WY was 4 666 (S=516) and 3 875 (S=305) kg, respectively. Node 8 is divided into fourteenth and fifteenth child-nodes as Node 14 and Node 15 in terms of region. Çermik, Eğil, Yenişehir and Bismil districts located in Node 14 and the average WY was 3 465 kg/da (S=437). Sur and Kayapınar districts located at Node 15, and the average WY was 2 085 kg/da (S=603). Node 10 was divided into sixteenth and seventeenth child-nodes as Node 16 and Node 17 based on IT. Ten producers in Node 16 among all watermelon producers, irrigated 2 hours or less obtained the highest WY yield (6492 kg/da). The results revealed that the irrigation is needed to obtain high WY, and the most appropriate amount of water should be given in irrigation.

CONCLUSION

Data mining method is not adequately known and is not widely used in scientific studies related to crop production of agricultural activities. Investigated factors affecting the yield in crop production and the predictions carried out with the algorithms in the regression tree analysis method can be effective in crop production. The watermelon yield obtained in the study area and in the year of the study (4 488 kg/da) was found above the World (3 262 kg/da) and Turkey (4 404 kg/da) average (FAO, 2019). The results revealed that watermelon producers in Diyarbakir should apply 5-6 hoeing, and 5-6 irrigation less than 2 hours, and harvest watermelon in June. Estimating the factors affecting the yield with the regression tree algorithms in the crop productions planned in the future will make a great contribution to scientific research in theory and practice.

Conflict of Interest

The article authors declare that there is no conflict of interest between them.

Author's Contributions

The authors declare that they have contributed equally to the article.

REFERENCES

Altaş, S, (2015). Investigation of In Vivo and In Vitro Antioxidant Activities of Diyarbakır Watermelon. (Doctoral Thesis). Dicle University Institute of Science, Diyarbakır. (In Turkish).

Anonymous, (2005).Watermelon cultivation, T.R. Ministry of Agriculture and Forestry, Samsun Provincial Directorate of Agriculture and Forestry https://samsun.tarimorman.gov.tr/Belgeler/Yayinlar/Lifletlerimiz/s-13.pdf

Anonymous, (2019a). Diyarbakır Metropolitan Municipality, Geography, Climate, Population Data. https://www.diyarbakir.bel.tr/diyarbakir/genel-bilgiler/cografi-bilgiler.html (In Turkish). (Date of access: 10 May 2022).

Anonymous, (2019b). T.R. Ministry of Agriculture and Forestry, Diyarbakır Provincial Directorate of Agriculture and Forestry Crop Production Records. https://diyarbakir.tarimorman.gov.tr/ (In Turkish). (Date of access: 10 May 2022).

Anonymous, (2021). World Watermelon Production by Countries. https://www.atlasbig.com/tr/ulkelerin-karpuz-uretimi

Aytekin, İ., Eyduran, E., Karadas, K., Akşahan, R., Keskin, İ, (2018). Prediction of Fattening Final Live Weight from some Body Measurements and Fattening Period in Young Bulls of Crossbred and Exotic Breeds using MARS Data Mining Algorithm. R. Bras. Zootec., 50(1):189-195.

Ban, D., Ban, SG., Oplanic, M., Horvat, J., Novak, B., Zanic, K., Znidarcic, D, (2011). Growth and Yield Response of Watermelon to In-Row Plant Spacing and Mycorrhiza. Chilean Journal of Agricultural Research 71(4):497-502.

Bostancı, B., Eren Atay, C, (2018). Decision Support Tools for Barley Yield: The Case of Menemen – Turkey. Dokuz Eylul University Faculty of Engineering Science and Engineering Journal, 20(60): 1057-1067.

Büyükkalay, H., 2019. Watermelon Production and Marketing Structure in Antalya Province. Antalya University, Institute of Science, Master's Thesis, page, 71.

Çat, A., Yardımcı, N., Kılıç, HÇ, 2016. Determination of Viral Factors in Greenhouse Cucumber (Cucumis sativus L.) and Cabbage (Cucurbita pepo L.) Production Areas in Antalya Province and Districts. Süleyman Demirel University Journal of the Institute of Science. 1, 129-132. (In Turkish)

Celik, S., Eyduran, E., Karadas, K., Tariq, MM., (2017). Comparison of predictive performance of data mining algorithms in predicting body weight in Mengali rams of Pakistan. Revista Brasileira de Zootecnia, 46(11): 863-872.

Celik, S., Eyduran, E., Tatliyer, A., Karadas, K., Kara, MK., Waheed, A, (2018). Comparing Predictive Performances of some Nonlinear Functions and Multivariate Adaptive Regression Splines (MARS) for Describing the Growth of Daera Dın Panah (DDP) Goat in Pakistan. Pakistan J. Zool., 50(3):1-4.

Duru, M., Duru, A., Karadas, K., Eyduran, E., Cinli, H., Tariq, MM, (2017). Effect of carrot (Daucus carota) leaf powder on external and ınternal egg characteristics of hy-line white laying hens. Pakistan Journal of zoology, 49: 125-132. https://doi. org/10.17582/journal.pjz/2017.49.1.125.132.

Eyduran, E., Zaborski, D., Waheed, A., Celik, S., Karadas, K., Grzesiak, W, (2017). Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the ındigenous beetal goat of Pakistan. Pakistan Journal of zoology, 49: 257-265. https:// doi.org/10.17582/journal.pjz/2017.49.1.257.265.

FAO, (2019). Food and Agricultural Organization of the United Nations. Crops and Livestock Products. https://www.fao.org/faostat/en/#data/QCL (Date of access: 16 May 2022).

Filho, FSO., Pereira, FHF., Brito, MEB., Medeiros, JE., Lacerda, F. H. D., Junior, JE C, (2019). Yield, quality and nutrient accumulation in watermelon as a function of organo-mineral fertilization. Comunicata Scientiae 10(1): 141-149.

Geçioğlu Erincik, B, (2015). The prevalence and intensity of watermelon fusarium wilt disease in Aydın, the breeds of the causative Fusarium oxysporum f. Sp. niveum (Fon), vegetative compatibility groups and the reactions of some watermelon varieties to the agent. (Doctoral Thesis). Adnan Menderes University Institute of Science, Aydın. (In Turkish).

Grzesiak, W., Zaborski, D, (2012). Examples of the use of data mining methods in animal breeding. Data mining applications in engineering and medicine, 303-324.

Güner, N. Wehner, T.C, (2004). The genes of watermelon. Hortscience, 39(6): 1175- 1182.

Güneş, R Aşkın, B, 2016). Chemical properties and nutritional content of watermelon seed oil. Food, 41(1): 37-44. (In Turkish)

Güngör, B, Balkaya, A., (2015). Mini Watermelon Cultivation. Journal of Turkish Seed Growers Association. 2: 26-29.

Irmak, S., Ercan, U, (2017). Determination of Factors Affecting Vegetable Oil Consumption with Data Mining Methods. Kafkas University Journal of Economics and Administrative Sciences Faculty, 8(15): 57-79. (In Turkish).

Karadas, K. and Kadirhanogullari, İH, (2017). Predicting Honey Production using Data Mining and Artificial Neural Network Algorithms in Apiculture. Pakistan Journal of zoology 49 (5):1611-1619.

Karadas, K., Birinci, A, (2019). Determination of factors affecting dairy cattle: a case study of Ardahan province using data mining algorithms. Revista Brasileira de Zootecnia, 48:1-11.

Karadas, K., Tariq, M., Tariq, MM, Eyduran, E. (2017). Measuring Predictive Performance of Data Mining and Artificial Neural Network Algorithms for Predicting Lactation Milk Yield in Indigenous Akkaraman Sheep. Pakistan Journal of zoology, 49(1):1-7.

Karakaya, E., Çelik, Ş., ve Taysı, MR, (2018). Investigation of Factors Affecting Fish Meat Consumption with CHAID Algorithm. Gaziosmanpaşa University Journal of Agriculture Faculty, 35 (2), 85-93. (In Turkish).

Kavut, YT., Geren, H., Simiã, A, (2014). Effect of Different Plant Densities on The Fruit Yield and Some Related Parameters and Storage Losses of Fodder Watermelon (Citrillus lanatus var. Cit). Turkish Journal of Field Crops, 19(2), 226-230.

Koç, Y., Eyduran, E., Akbulut, O, (2017). Application of Regression Tree Method for Different Data from Animal Science. Pakistan Journal of zoology, 49(2): 599-607.

Koçkaya, MA, 2019. Effects of Different Fertilizer Types on Yield and Quality in Diyarbakır Watermelon (Citrullus Lanatus). (Master Thesis). Dicle University Institute of Science, Diyarbakır. (In Turkish).

Koleboshina, T. G., Varivoda, E. A. 2020. Melon growing industry analysis in modern economic conditions. In IOP Conference Series: Earth and Environmental Science (Vol. 459, No. 6, p. 062075). IOP Publishing. april

Küçükönder, H., Vursavuş, KK., Üçkardeş, F, 2015. Determining the Effect of Some Mechanical Properties on Color Maturity of Tomatoes by K-Star, Random Forest and Decision Tree (C4.5) Classification Algorithms. Turkish Journal of Agriculture Food Science and Technology, 3(5): 300-306. (in Turkish).

Kuşcu, H., Turan, A., Özmen, N., Aydınol, P., Demir, AO., 2015. Effects of different irrigation regimes on water use efficiency, yield and fruit quality of watermelon under Bursa ecological conditions. Journal of Akdeniz University Faculty of Agriculture, 28(1): 21-26.

Malueva, S. V., Bocherova, I. N., & Kornilova, M. S. 2020. Use of the source material in the selection of watermelon and melon. News of FSVC, (2), 68-72.

Oktay, A., Doran, I. 2005. The Effects of Nitrogen Fertilization on Fruit Yield and Quality of Turkey's Largest Watermelon Spreading Variety. Journal of Akdeniz University Faculty of Agriculture, 18(3), 305-311.

Okumuş, V, 2010. Biotechnological Research on the Micropropagation of Diyarbakır Watermelon Genotypes (Citrullus Lanatus Cv. 'White Winter', 'Karakış' and 'Spread'). (Doctoral Thesis). Dicle University Institute of Science, Diyarbakır. (In Turkish).

Özçınar, S, 2020. A Characterization of Watermelon Mosaic Virus (Watermelon Mosaic Virus, Wmv) in Watermelon and Melon Fields of Calf and Mersin Provinces. (Master Thesis). Çukurova University Institute of Science, Adana. (In Turkish)

Öztürk, N, 2018. Quantitative Real Time PCR Diagnosis and Detection of Watermelon Bacterial Fruit Spot Disease Acidovorax Citrulli and Investigation of Sensitivity Reactions of Watermelon and Melon Varieties. (Doctoral Thesis) Akdeniz University Institute of Science and Technology, Antalya. (In Turkish).

Pejic, B., Mackic, k., Pavkovic, S., Lejevnaic-Masic, B., Aksic, M., Gvozdanovic-Varga, J, 2016. Water-Yield Relations of Drip Irrigated Watermelon in Temperate Climatic Conditions. Contemporary Agriculture 65(1):53-59.

Rolbiecki, R., Rolbiecki, S., Piszczek, P., Figas, A., Jagosz, B., Ptach, W., Prus, P., Kazula, M. J, (2020). Impact of Nitrogen Fertigation on Watermelon Yield Grown on the Very Light Soil in Poland. Agronomy, 10:1-10.

Şanlı, A., Kaya, M., Kara, B, (2009). Effects of Weed Control Times and Herbicide Applications on Yield and Some Yield Components in Chickpea (Cicer arietinum L.). Anatolian Journal of Agricultural Sciences, 24(1), 13-20. (In Turkish).

Seçer, A., Çelik, F., Barut, H., (2020). The Factors Affecting the Producers' Decision to Grow Fruit Tree and Their Expectations for the Future in Ağrı Province. Çukurova Journal of Agriculture and Food Science, 35(2): 77-88.

Şimşek, A., Dinler, H., Duru, S., (2020). Determination of Approaches of the Stone Fruit Producers to the Phytopathological Problems of Uşak Province. International Journal of Life Sciences and Biotechnology, 3(2): 127- 147.

Sun, J., Hui, LI, 2008. Data Mining Method for Listed Companies, Financial Distress Prediction. Knowledge-Based Systems, 21, No. 1.

Sylvestre, H., Bosco, N. J., Emmanuel, N., Christine, U, 2014. Growth and yield of Watermelon as affected by different spacing and mulching types under Rubona conditions in Rwanda. Scholarly Journal of Agricultural Science, 4(10): 517-520.

Tatlıyer, A, 2020. The Effect of Breeding Type on the Prediction Performance of Regression Tree Algorithms in Lambs. Kahramanmaraş Sütçü İmam University Journal of Agriculture and Nature,23(3), 772-780. (In Turkish).

Tokgöz, H., Gölükcü, M., Toker, R., Turgut, DY, 2015. The Effects of Grafted Seedling Use and Harvest Time on Some Physical and Chemical Properties of Watermelon (Citrullus Lanatus), Gıda, 40(5), 263-270. (In Turkish).

TUİK, 2021. Turkish Statistical Institute. https://biruni.tuik.gov.tr/medas/?kn=92&locale=tr January 21, 2021. (In Turkish).

Tuna, A. L., Ozer, O. 2005. Effect of Different Calcium Compounds on the Fruit Yield, Nutrition and some Quality Properties of Watermelon (Citrullus lanatus) Plant. Journal of The Faculty of Agriculture, 42(1), 203.

Vural, Ç., Dağdelen, N, 2008. The Effects of Different Irrigation Programs on Yield and Some Agronomic Properties of Popcorn Irrigated by Drip Irrigation Method. Adnan Menderes University Journal of the Faculty of Agriculture, 5(2), 97-104. (In Turkish)

Wehner, TC, 2010. Watermelon crop information. North Carolina State University. Raleigh, NC.

Yamane, T, 2010. Basic Sampling Methods. Gazi University Faculty of Science and Letters, Department of Statistics, Literature Publications, No.53, 116 p., Istanbul (In Turkish)

Yavuz, D., Seymen, M., Süheri, S., Yavuz, N., Türkmen, Ö., & Kurtar, E. S. 2020. How do rootstocks of citron watermelon (Citrullus lanatus var. citroides) affect the yield and quality of watermelon under deficit irrigation?. Agricultural Water Management, 241, 106351.

Zaborski, D., Ali, M., Eyduran, E., Grzesiak, W., Tariq, M. M., Abbas, F., Waheed, A., Tirink, C, 2019. Prediction of Selected Reproductive Traits of Indigenous Harnai Sheep under the Farm Management System via various Data Mining Algorithms. Pakistan Journal of zoology, 51(2): 421-431.