Atatürk Üniv. Ziraat Fak. Derg., 49 (1): 37-43 , 2018
Atatürk Univ., J. of the Agricultural Faculty, 49 (1): 37-43 , 2018
ISSN : 1300-9036 Araştırma Makalesi/Research Article

Data Mining Approach For Prediction Of Fruit Color Properties

Bünyamin DEMİR[1*] Feyza GÜRBÜZ[2] İkbal ESKI[3] Zeynel Abidin KUŞ[4]
[1/]Mersin University, Vocational School of Technical Sciences, Department of Mechanical and Metal Technologies, Mersin, Turkey
[2/:]Erciyes University, Faculty of Engineering, Department of Industrial Engineering, 38039, Kayseri, Turkey
[3/:]Erciyes University, Faculty of Engineering, Department of Mechatronics Engineering, 38039, Kayseri, Turkey
[4/:] Erciyes University, Faculty of Agriculture, Department of Biosystems Engineering, 38039, Kayseri, Turkey
(*Corresponding author email: bd@mersin.edu.tr)

Geliş Tarihi :13.12.2017 Kabul Tarihi :08.01.2018

ABSTRACT: Color is an important feature that dictates the quality and consumer preferences of many fresh fruits and vegetables. In color measurement of fruits, the CIE L*a*b* color space is widely used since it is a uniform color scale. In this study, raw data for the color features of apple varieties were divided into two parts as test and train data in the first stage, analyses were performed on train data and tests were performed on test data. The rules obtained by applying the Find laws algorithm were used to estimate the color index (CI), hue angle (h *) and Chroma (C *) values. In the second stage, raw data were classified by Strict and Liberal options of cluster analysis. Find Laws algorithm was applied to each cluster and 7 different prediction rules were obtained for CI, h*and C* parameters. R^2 values of the rules were compared and the rules with the most accurate outcomes were identified.

Keywords: Apple, hue angle, L*a*b*, color space.

Meyve Renk Özelliklerini Tahmin Etmek İçin Veri Madenciliği Yaklaşımı

ÖZET: Renk, birçok taze meyve ve sebzenin kalitesini ve tüketici tercihlerini belirleyen önemli bir özelliktir. Meyvelerin renk ölçümünde, uniform renk ölçeği nedeniyle CIE L*a*b* en çok kullanılan renk uzayıdır. Bu çalışmada elma çeşitlerinin renk özelliklerine ait ham veriler ilk aşamada test ve eğitim verileri olarak iki kısma ayrılmış, eğitim verileri üzerinde analizler yapılmış ve test verileri ise testlerde kullanılmıştır. Find laws algoritması uygulanarak elde edilen kurallar Color index (CI), hue angle (h*) and Chroma (C*) değerlerini tahmin etmek için kullanılmıştır. İkinci aşamada ise ham veriler cluster analizine tabi tutularak Strict ve Liberal seçenekleri ile sınıflandırılmıştır. Find laws algoritması her bir sınıfa tek tek uygulanıp, her bir CI, h*, C* parametreleri için elde edilen 7 farklı tahmin kuralı R^2 değerlerine göre karşılaştırılarak en yüksek doğruluğa sahip kurallar tespit edilmiştir.

Anahtar kelimeler: Elma, hue açısı, L*a*b*, renk uzayı

INTRODUCTION

Visual appearance is the primary quality attribute for foodstuffs and it is also the first quality aspect looked for by the consumers (Maskan, 2001). For fruits, size, shape, texture, color and surface defects are the basic external quality attributes. All these attributes are also related to visual appearance of the fruits and agricultural products (Zhang *et al.*, 2017). Among these visual quality attributes, color is the most significant parameters used as an indicator of the quality. Therefore, majority of the consumers first look at the color of fruits, vegetables and meats to judge the quality of those foodstuffs (Wu and Sun, 2013; Trinderup *at al.*, 2015).

Peel color is the key quality attribute for apples. It not only influences consumer preferences, but also is related to nutritional values of the apples. Peel color is also used to distinguish one cultivar from the other since each cultivar has specific color characteristics (Rabinovich, 2009). Multiple color spaces are often used to define color parameters of fruits. Among them, CIE L*a*b* (CIELAB) specified by International Commission on Illumination is the most common one (Fairchild, 2013).

Data mining approach uses various tools and techniques to inquire meaningful data from a large data set. It has recently started to be used in agricultural researches and implementations (Chowdhury and Ojha, 2017). Several previous researches employed data mining techniques in agricultural researches to predict the value of an attribute by using already available measurements of that attribute (Ramesh and Vardhan, 2013; Gonzalez-Sanchez *et al.*, 2014; Veenadhari *et al.*, 2014; Pantazi *et al.*, 2016; Germšek *et al.*, 2017; Isaza *et al.*, 2017; Kus *et al.*, 2017; Majumdar *et al.*, 2017).

Data mining approach includes four basic methodologies as of clustering, classification, feature selection and outlier detection. On the other hand, artificial neural networks, decision trees, k-means type algorithms, genetic algorithms, nearest neighbor method and rule induction are the primary techniques

used in data mining. Such techniques have successfully been implemented for fruits and vegetables mostly focusing on classification of apples, citrus, strawberries, table olives, onions (Vlontzos and Pardalos, 2017). Apple is the most widely consumed fruit worldwide. It is consumed either as fresh or processed. It has the second place in world fruit production (Cebulj *et al.,* 2017; Ortiz *et al.,* 2017). Therefore, apple was selected as the research material of this study.

The primary objective of the present study was to estimate color parameters of apples with different color characteristics by using data mining approaches.

MATERIAL AND METHODS

Six apple varieties, Amasya, Arapkızı, Golden Delicious, Granny Smith, Starking, Pink Lady, were used to evaluate the skin color parameters (L*, a*, and b*). L* represents lightness (100: white, 0: black), a* indicates the difference between red (+a*) and green (-a*), and b* represents the difference between yellow (+b*) and blue (-b*). Fifty fruits were selected from each apple varieties and the skin color of fruits was measured over the cheek areas with a Minolta Chroma Meter CR-400 (Minolta-Konica, Japan) on the basis of CIE L*a*b* color space. Four measurements were taken at the equator lateral section of fruit through rotating the apple 90° between each acquisition. Color index (CI), hue angle (h*) and Chroma (C*) values were derived from a* and b* values. Equations in the scientific literatures (McGuire, 1992; McLellan *et al.,* 1995; Viscarra Rossel *et al.* 2006; Kus *et al.,* 2017) were used for calculations.

Cluster Analysis

Data mining covers all the methods applied in data analysis techniques to find out earlier unrecognized valid samples and relationships in huge datasets. Some examples of these techniques include classification, data summarization, anomaly detection, dependency finding, regression and clustering (Han and Kamber, 2000).

Partitioning of data points into several set of groups on the basis of similarity between data points is called clustering (Jain and Dubes, 1988). These groups are named as clusters (Ahmad and Hashmi, 2016). Clustering is widely used in several fields, such as data mining, knowledge discovery, machine learning, statistics and includes the roots of data clustering (Cheng *et al.,* 2013; Kao *et al.,* 2008; Leung *et al.,* 2000; Nguyen and Cios, 2008; Sahoo *et al.,* 2012; Thong *et al.,* 2015; Qiu *et al.,* 2016; Saha *et al.,* 2016).

Clustering is a quite significant issue especially in rapidly growing fields such as knowledge discovery and data mining (Armano and Farmani, 2016). Therefore, Cluster Analysis engine of PolyAnalyst was employed in this study.

Cluster Analysis in PolyAnalyst

The Cluster Analysis engine examines a dataset for areas of similarity. The datasets are compared for all attributes and similarities and differences are found. The use of all attributes makes the Cluster algorithm very useful for beginning data mining. It is used for location of anomalies in data, undirected data mining – discovery of unknown relationships in data and preprocessing – division of data into groups of similar records for further analysis. It doesn't require any target attribute and will work on any data. Outputs of Cluster analysis are datasets containing clusters found and prediction table. Clustering is one of PolyAnalyst's preprocessing methods (PolyAnalyst, 2007).

PolyAnalyst's clustering algorithm, LA (Localization of Anomalies), searches for only significant clusters of data, it needs a sufficient amount of data in order to guarantee that the obtained result is more than merely an accidental fluctuation. LA algorithm can select several attributes, most important for clustering, from all attributes of the explored dataset (PolyAnalyst, 2007).

There are two modifications (Strict and Liberal) of Cluster analysis. The Liberal algorithm will find larger clusters and place much of the dataset into clusters, while the Strict setting will only find true anomalies in the data. Both of them were used in this study as preprocessing stages (PolyAnalyst, 2007).

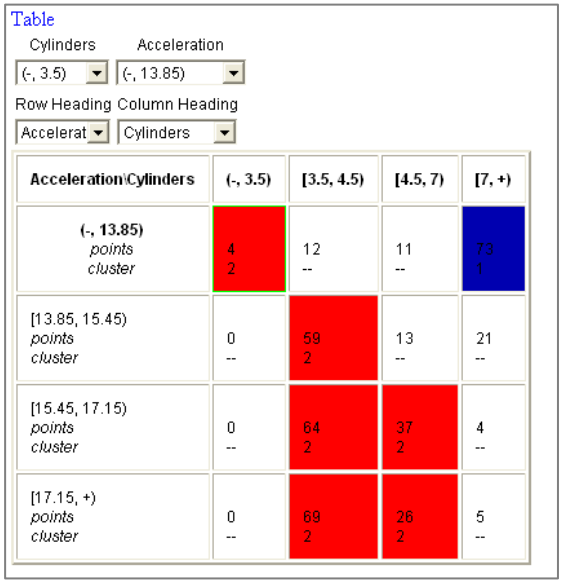
The report of cluster analysis begins with listing the "Set of parameters that give the best clustering". This is a list of the attributes that were used to divide the data into clusters. Generally, only a few attributes are used. Next, the p-values of clustering is listed – lower values indicate a higher degree of significance. The "Total N of all points in all clusters" indicates how many of the data records were classified into clusters–generally, this is substantially less than the number of records in the dataset. Finally, the number of clusters are found and the number of data points in all clusters is listed. An example of this report is presented in Fig.1 (PolyAnalyst, 2007).

Set of parameters which give the best clustering: Cylinders Acceleration	
P-value:	1.316e-030
Total N of points in all clusters:	332
Number of clusters found:	2

Fig. 1. A sample text report of PolyAnalyst Cluster Analysis.

After the text report shown in Fig. 1, a clustering table is displayed as given in Fig. 2. This

table divides the data into segments along two axes – the two axes being the attributes that had the strongest effect on clustering by default. Each cell is color-coded depending on which cluster it falls into – most cells are left white, as they do not fall into any clusters. Within each cell, two numbers are listed – the number of points in that cell and the cluster number for that cell (PolyAnalyst, 2007).



Table

Cylinders Acceleration

(-, 3.5) (-, 13.85)

Row Heading Column Heading

Accelerat Cylinders

Acceleration\Cylinders	(-, 3.5)	[3.5, 4.5)	[4.5, 7)	[7, +)
(-, 13.85) *points* *cluster*	4 2	12 --	11 --	73 1
[13.85, 15.45) *points* *cluster*	0 --	59 2	13 --	21 --
[15.45, 17.15) *points* *cluster*	0 --	64 2	37 2	4 --
[17.15, +) *points* *cluster*	0 --	69 2	26 2	5 --

Fig. 2. An example of table report of PolyAnalyst Cluster Analysis.

Find Laws

Find Laws exploration engine is one of the most powerful and absolutely unique algorithms for data exploration implemented in PolyAnalyst. Its purpose is an automated discovery of multi-dimensional nonlinear connections in the data and presentation of these connections in the form of explicit mathematical notation, which can be readily understood and analyzed by the user. The data analysis techniques used in this method are based on automated synthesis of functional programs treated as multi-dimensional non-linear regression models. The internal programming language has a sufficient expressive power to formalize any relation which can be expressed in an algorithmic form. The main advantage of this approach is its ability to discover in the data hidden relations, which might assume a great variety of forms (PolyAnalyst, 2007).

Find laws is generally used in the final process of data mining application to present a human readable rule explaining the analysis. R-squared is the measure of the accuracy and efficiency of a built model. R-squared is equal to $1 - \rho^2$, where ρ is the standard error. Its values also lie in the interval [0,1], but in contrast with the standard error, it is equal to 1 in the case of an absolutely accurate model and to 0 – when the mean value of the target variable is taken as a prediction for all records. R-squared can be roughly interpreted as a part of the target parameter variability explained by the discovered model (PolyAnalyst, 2007, Gürbüz *et al.*, 2011).

RESULTS AND DISCUSSION

The summary of this study is shown graphically below in Fig. 3.

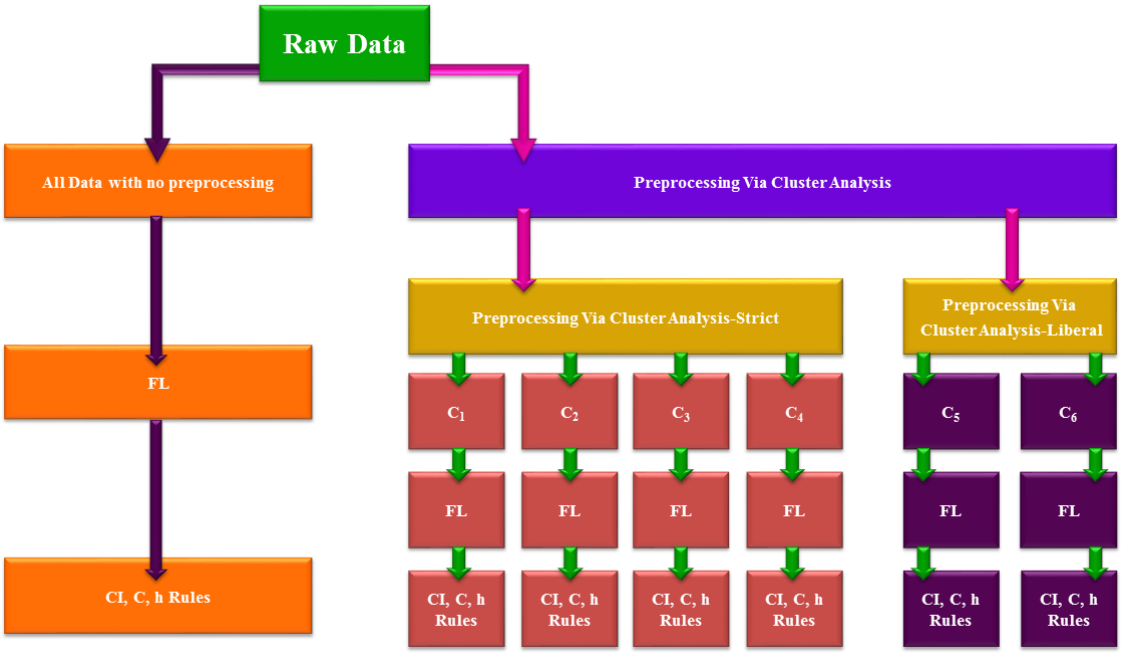


Fig. 3. The graphical summary of this study.

As seen in Fig. 3, firstly, Find Laws engine was applied to have prediction rules without preprocessing on raw data. Then, Cluster Analysis was applied to find anomalies and inefficient data in raw data set. As stated earlier, Cluster Analysis engine of PolyAnalyst has two modifications as of Strict and Liberal. Firstly, the Strict version was applied and four clusters were obtained. As seen in Table 1 and Fig. 4, the report lists "CI" and "C*" attributes that were used for developing the clusters. The total number of points entered into each cluster (306 data points out of 600 belong to one of the four clusters) and the number of detected clusters are presented in Figure 4. Each cell in the table contains points which belong to different clusters, and is marked by a different color. The rest of the data points, which have a more homogeneous distribution, are not colored. Then, Liberal version was applied to the same raw data and two clusters were observed as given in Fig. 5. As it was in Strict version, the report lists "CI" and "C*" attributes that were used for developing the clusters (Table 2). The total number of points entered into each cluster (384 data points out of 600 belonged to one of two clusters). The specifications for all clusters are also shown in Fig. 4 and Fig. 5 depending on "C*" and "CI".

Clustering creates databases which correspond to clusters. In this case, four new datasets (C_1, C_2, C_3, C_4) were observed after Cluster Analysis-Strict version and two new datasets (C_5, C_6) after Cluster Analysis-Liberal version.

Table 1. The text results of Cluster Analysis-Strict.

Set of parameters which give the best clustering	**CI, C***
P-value	8.08e-024
Total N of points in all clusters	306
Number of clusters found	4

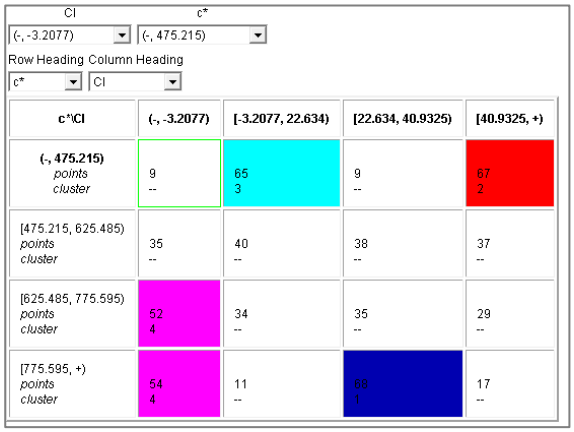


Fig. 4. The specifications of all clusters found by Cluster Analysis-Strict.

Table 2. The text results of Cluster Analysis Liberal.

Set of parameters which give the best clustering	CI, C*
Total N of points in all clusters	384
Number of clusters found	2

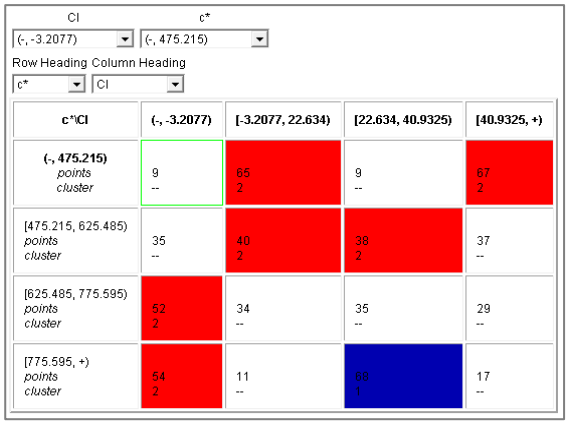


Fig. 5. The specifications of all clusters found by Cluster Analysis-Liberal.

After preprocessing via cluster analysis strict and liberal version, there were six clusters to be analyzed via Find Laws. The Find Laws was applied to raw data and six clusters to have prediction rules for the outputs (CI, C*, h*) (C_1, C_2, C_3, C_4, C_5, C_6) separately.

The rules obtained via raw data are shown in Table 3. In Tables 4-9, the results and the rules obtained by six clusters (C_1, C_2, C_3, C_4, C_5, C_6) are given separately.

Table 3. The rules obtained via raw data.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = 802613 *a/(1+802.61 *b*L)	1	1
C*	C* = (0.0248452 *b*h*$a^{6)}$ +2563.09 *b*a^4-300.449 *h*a^4+2.95208e+006 *a^2+0.00845857 *h*a^7+31916.5 *a^3-6.10791e^{+007} *a+678907 *h*a)/(b*a^4+66.864 *a^4-120.373 *h*a^2+25814.2 *a^2+40584.5)	1	1
h*	h* = (5.63601 *CI*CI+90.0291 *CI+0.0220817 *CI*CI*CI-1.10645e^{-005} *c*c*CI-0.00111407 *c*CI*CI-3.74551e^{-005} *c*c)/(CI+0.100361 *CI*CI+0.0035753 *CI*CI*CI)	0.9996	0.999966

Table 4. The rules obtained via C_1 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = (1000 *a*a-37081.9 *a)/(a*b*L-37.0818 *b*L)	1	0.999974
C*	C* = (-0.00478807 *a+1.00041 *b*a)/(1+8.81289e^{-006} *b)	1	0.927550
h*	h* = (77.8128 *b^4*a-10.8865 *b^5-6.32136 *b^3*a^2+0.769582 *c^2*a)/(b^3*a^2+0.36419 *c*b^3)	1	0.874395

Table 5. The rules obtained via C_2 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = (57311.2 -5.47176 *h*h-108.527 *h)/(h*L-0.037002 -0.00181481 *h*h*L)	1	0.997706
C*	C* = 0.00021626 *b+0.999995 *a*b	1	0.827664
h*	h* = (17.9313 *b^3*a-0.518119 *b^2*a^2+120.398 *b*a^3+2.48383 *b^4-3.59334 *b^2+26.4973 *b)/(b^2*a^2+0.0111565 *b^3*a+2.10002 *a^4)	1	0.787990

Table 6. The rules obtained via C_3 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = (-0.161087 *L*a-4.45967 *L+62841.5 -10.6112 *h*h-0.000645239 *h^4+0.0902219 *h^3)/(L*h+72.9961)	1	0.999504
C*	C* = (-0.19737 *a^4+659.699 *a^2)/(a*a+1.61315 *a+63.6683)	0.9502	0.821465
h*	h* = (14809.1 *b^2-169.549 *a*b^2-1887.45 *a*b-612.005 *b-19.0871 *a^3+106.224 *a^2*b+12526.3 +0.130525 *a3*b+1006.65 *a^2)/(b-1.89046 *a*b^2+83.9796 *"a*"*"b*"+164.437 *b*b+63.6255 *a^2)	1	0.999830

Table 7. The rules obtained via C_4 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = (0.00016783 *L*a+999.989 *a)/(L*b)	1	1
C*	C* = (-0.00406219 *a-1.00011 *b*a-0.0795031)/(1-5.77947e^{-006} *a)	1	0.721373
h*	h* = (-0.000670031 *CI*CI+89.3236 *CI-0.0613815 *L*CI*CI+0.00554636 *L-2.27983)/(CI+1.20232e^{-007} *L^2*CI*CI*CI-0.0132279)	1	0.306836

Table 8. The rules obtained via C_5 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = (1000 *a^2-37081.9 *a)/(a*b*L-37.0818 *b*L)	1	0.999974
C*	C* = (-0.00478807 *a+1.00041 *b*a)/(1+8.81289e^{-006} *b)	1	0.827550
h*	h* = (65.5363 *b*c*a^2+7.44529 *b*a^4-75.8628 *b*c^2+17.1324 *b^3*c)/(b*a^4-0.831844 *b*c*a^2+0.14819 *a^5)	1	0.897578

Table 9. The rules obtained via C_6 dataset.

Output	Prediction Rule	R^2 Train	R^2 Test
CI	CI = 6.76534e+006 *a/(b+6765.32 *L*b)	1	1
C*	C* = (-0.000229756 *if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)*if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)*if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)-1.00005 *b*a*if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)*if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)+1.03591 *b*a)/(if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)*if(-0.000601536 <= 1 /a and 1 /a < -0.000601536 + 1.25308 ,1 ,-3.33334 *a)+3.10548e-009 *a-0.964136)	1	1
h*	h* = (1.79074 *CI*b^2+18.1835 *CI*b+471.459 *CI-23.6002 *b+861.547)/(CI*b+0.0378674 *CI*CI*b+9.82629e^{-005} *CI*b^3)	0.9985	0.996466

CONCLUSION

In this study, Find Laws analysis on raw data yielded rules for CI, C* and h* parameters (Table 3). For Cluster Analysis, two modifications (Strict and Liberal on raw data. Cluster analysis yielded six clusters, four of which by Strict version (C_1, C_2, C_3, C_4) and two by Liberal version (C_5, C_6) depending on CI and C* parameters. Find Laws yielded new data subsets and the rules for these subsets. For example, for cluster C_1, there were 68 records that obeyed the rules obtained by this subset and depending on "C*" and "CI", features of this cluster records were "C* ≥ 775.595" and "22.634 ≤ CI < 40.9325" as seen in Table 10. It means that the obtained rules were significant and efficient for apples (C* ≥ 775.595" and "22.634 ≤ CI < 40.93259. In Table 10, number of records and features of all clusters are provided.

All R^2 values of train and test results were acceptable as given above in Tables 3-10. Most of "R^2 Train" values were equal to 1. It means that all the rules were efficient and acceptable for this apple datasets. And also "R^2 Test" and "R^2 Train" values were close to each other. This shows that the obtained rules were accurate for predicting output parameters. If the cluster features are to be specified, cluster analysis can be used as preprocessing; otherwise the raw data results should be used.

As seen in Table 10, the features of C_1 and C_5 are the same. And also some features of C_6 intersect with the clusters obtained by Cluster analysis-Strict version and also the rules obtained by C_1.

C_5 datasets were the same as seen in Table 4 and Table 8 for "CI" and "C*". For "h*" parameter, they are nearly the same because of the test results. As a result, it can be stated that Cluster analysis-Strict version results covered the Liberal version results as seen in Table 10. So applying only one of them can be enough for next studies.

Table 10. Number of records and features of clusters.

Cluster	Number of records	Features
C_1	68	(C* ≥ 775.595) and (22.634 ≤ CI < 40.9325)
C_2	67	(C* < 475.215) and (CI ≥ 40.9325)
C_3	65	(C* < 475.215) and (3.2077 ≤ CI < 22.634)
C_4	106	(625.485 ≤ C* < 775.595) and (CI < 3.2077) (C* ≥ 775.595) and (CI < 3.2077)
C_5	68	(C* ≥ 775.595) and (22.634 ≤ CI < 40.9325)
C_6	316	(C* < 475.215) and (3.2077 ≤ CI < 22.634) (475.215 ≤ C* < 625.485) and (3.2077 ≤ CI < 22.634) (475.215 ≤ C* < 625.485) and (22.634 ≤ CI < 40.9325) (625.485 ≤ C* < 775.595) and (CI < 3.2077) (C* ≥ 775.595) and (CI < 3.2077) (C* < 475.215) and (CI ≥ 40.9325)

REFERENCES

Ahmad, A., 2016. Sarosh Hashmi, K-Harmonic means type clustering algorithm for mixed datasets, Applied Soft Computing 48 39–49.

Armano, G., Farmani M.R. 2016. Multiobjective clustering analysis using particle swarm optimization, Expert Systems With Applications 55, 184–193.

Cebulj, A., Cunja, V., Mikulic-Petkovsek, M., Veberic, R. 2017. Importance of metabolite distribution in apple fruit. Scientia Horticulturae, 214, 214-220.

Cheng, H., Yang, S., Cao, J., 2013. Dynamic genetic algorithms for the dynamic load balanced clustering problem in mobile ad hoc networks. Expert Systems with Applications, 40 (4), 1381–1392.

Chowdhury, D.R., Ojha, S., 2017. An Empirical Study on Mushroom Disease Diagnosis: A Data Mining Approach. International Research Journal of Engineering and Technology (IRJET), 04(01), 529-534

Fairchild, M. D. 2013. Color appearance models. John Wiley & Sons.

Germšek, B., Rozman, Č., Unuk, T., 2017. Forecasting Apple Fruit Color Intensity with Machine Learning Methods. Erwerbs-Obstbau, 59(2), 109-118.

Gonzalez-Sanchez, A., Frausto-Solis, J., Ojeda-Bustamante, W., 2014. Predictive ability of machine learning methods fo r massive crop yield prediction. Spanish Journal of Agricultural Research, 12(2), 313-328.

Gürbüz, F., Özbakır, L., Yapıcı, H., 2011. Data mining and preprocessing application on component reports of an airline company in Turkey. Expert Systems with Applications, 38, 6618–6626.

Han, J., Kamber, M., 2000. Data mining: concepts and techniques, the Morgan Kaufmann Series in data management systems. Morgan Kaufmann.

Isaza, C., Anaya, K., de Paz, J.Z., Vasco-Leal, J.F., Hernandez-Rios, I., Mosquera-Artamonov, J.D., 2017. Image analysis and data mining techniques for classification of morphological and color features for seeds of the wild castor oil plant (Ricinus communis L.). Multimedia Tools and Applications, 1-18.

Jain, A.K., Dubes R.C., 1988. Algorithms for Clustering Data, Prentice-Hall, Inc., 1988.

Kao, Y.T., Zahara, E., Kao, I.W, 2008. A hybridized approach to data clustering. Expert Systems with Applications, 34 (3), 1754–1762. doi: 10.1016/j.eswa.2007. 01.028.

Kus, Z.A., Demir, B., Eski, I., Gurbuz, F., & Ercisli, S., 2017. Estimation of the Colour Properties of Apples Varieties Using Neural Network. Erwerbs-Obstbau, 59(4), 291-299.

Leung, Y., Zhang, J.S., Xu, Z.B., 2000. Clustering by scale-space filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (12), 1396–1410. doi: 10.1109/34.895974 .

Majumdar, J., Naraseeyappa, S., Ankalaki, S., 2017. Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data, 4(1), 20.

Maskan, M., 2001. Kinetics of colour change of kiwifruits during hot air and microwave drying. Journal of food engineering, 48(2), 169-175.

McGuire, R.G., 1992. Reporting of objective color measurements. HortScience 27(12): 1254-1255.

McLellan, M.R., Lind, L.R., Kime, R.W., 1995. Hue angle determination and statistical analysis for multiquadrant hunter L, a, b data. J Food Qual 18(3):235–240

Nguyen, C.D., Cios, K.J., 2008. Gakrem: a novel hybrid clustering algorithm. Information Sciences, 178 (22), 4205–4227. doi: 10.1016/j.ins.2008.07.016.

Ortiz, A., Le Meurlay, D., Lara, I., Symoneaux, R., Madieta, E., Mehinagic, E., 2017. The effects of sous-vide cooking parameters on texture and cell wall modifications in two apple cultivars: A response surface methodology approach. Food Science and Technology International, 23(2), 99-109.

Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R. L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. Computers and Electronics in Agriculture, 121, 57-65.

PolyAnalyst, 2007. User Manuel of PolyAnalyst 6.5, April 2007.

Qiu, H., Xu, Y., Gao, L., Li, X., Chi, L., 2016. Multi-stage design space reduction and metamodeling optimization method based on self-organizing maps and fuzzy clustering. Expert Systems with Applications, 46, 180–195.

Ramesh, D., Vardhan, B.V., 2013. Data mining techniques and applications to agricultural yield data. International Journal of Advanced Research in Computer and Communication Engineering, 2(9), 3477-80.

Rabinovich, A. T., 2009. Studies on apple peel color regulation. University of Minnesota.(pp.1-65)

Saha, S., Alok, A.K., Ekbal, A., 2016. Brain image segmentation using semi- supervised clustering. Expert Systems with Applications, 52, 50–63.

Sahoo, A.K., Zuo, M.J., Tiwari, M., 2012. A data clustering algorithm for stratified data partitioning in artificial neural network. Expert Systems with Applications, 39 (8), 7004–7014.

Thong, N.T., 2015. HIFCF: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. Expert Systems With Applications, 42(7), 3682–3701.

Trinderup, C. H., Dahl, A., Jensen, K., Carstensen, J.M., Conradsen, K., 2015. Comparison of a multispectral vision system and a colorimeter for the assessment of meat color. Meat science, 102, 1-7.

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006. Color space models for soil science. Geoderma 133:320–337

Vlontzos, G., Pardalos, P.M., 2017. Data mining and optimisation issues in the food industry. International Journal of Sustainable Agricultural Management and Informatics, 3(1), 44-64.

Veenadhari, S., Misra, B., Singh, C.D., 2014. Machine learning approach for forecasting crop yield based on climatic parameters. In Computer Communication and Informatics (ICCCI), 2014 International Conference on (pp. 1-5). IEEE.

Wu, D., Sun, D.W., 2013. Colour measurements by computer vision for food quality control–A review. Trends in Food Science & Technology, 29(1), 5-20.

Zhang, B., Dai, D., Huang, J., Zhou, J., Gui, Q., Dai, F., 2017. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review. Critical reviews in food science and nutrition, 1-20.