Dokuz Eylül Üniversitesi-Mühendislik Fakültesi
Fen ve Mühendislik Dergisi
Cilt 20, Sayı 60, Eylül, 2018

Dokuz Eylul University-Faculty of Engineering
Journal of Science and Engineering
Volume 20, Issue 60, September, 2018

DOI: 10.21205/deufmd. 2018206085

Decision Support Tools for Barley Yield: The Case of Menemen - Turkey

Büşra BOSTANCI[1], Canan EREN ATAY*[2]

Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, 35390 Izmir
[1](ORCİD:0000-0002-2196-4357)
[2](ORCİD:0000-0002-7706-7196)

(Alınış / Received: 18.05.2018, Kabul / Accepted: 20.06.2018,
Online Yayınlanma / Published Online: 15.09.2018)

Keywords
Decision Support System,
Data Mining,
Decision tree,
CHAID algorithm,
CR&T algorithm,
Agriculture

Abstract: The estimation of agricultural yield is a challenging and essential task for every farmer. Since the very old times, agriculture has always been the most important means of livelihood both in Turkey and all around the world. There are many factors that directly affect the efficiency in agriculture such as climatic features, use of water resources, proper and timely use of pesticides and fertilizers. Computer-based systems are needed to transform agriculture data into tangible information. Data mining involves certain methods of obtaining or inferring meaningful and otherwise-unknown information from the data. With the increasing significance of precision agricultural practices, farmers have become inclined to be engaged in a more conscious strategy of agriculture. In this study, barley crop data received from İzmir Menemen Provincial Directorate of Agriculture was carefully organized and evaluated with the classification algorithms in the SPSS Clementine software. CHAID and CR&T algorithms were employed and major factors that affect crop yield was defined. Based on these, a decision support system has been developed for farmers to forecast both harvest season and crop yield.

Arpa Verimi için Karar Destek Araçları. Menemen Örneği - Türkiye

Anahtar Kelimeler
Karar Destek Sistemleri,
Veri Madenciliği,
Karar ağacı,
CHAID algoritması,
CR&T algoritması,
Tarım

Özet: Tarımsal verimin tahmini, her çiftçi için zorlu ve önemli bir görevdir. Tarım geçmişten günümüze kadar hem Türkiye hem de bütün dünyada birçok insan için en önemli geçim kaynağı olmuştur. İklimsel özellikler, su kaynaklarının kullanımı, tarımsal ilaçlar ve gübrelerin doğru ve zamanında kullanılması gibi tarımın etkinliğini doğrudan etkileyen birçok faktör vardır. Tarımsal verileri anlamlı bilgiye dönüştürme sürecinde bilgisayar destekli sistemlere ihtiyaç vardır. Veri madenciliği, verilerden anlamlı ve başka türlü bilinmeyen bilgiler elde etmek veya çıkarmak için belirli yöntemler içerir. Hassas tarım uygulamalarının artan önemi ile çiftçiler daha bilinçli bir tarım stratejisi ile meşgul olmaya eğimli hale geldi. Bu çalışmada, İzmir Menemen Tarım il

müdürlüğünden alınan Arpa ürünün ekim verileri dikkatle düzenlenmiş ve SPSS Clementine yazılımındaki sınıflandırma algoritmaları ile değerlendirilmiştir. CHAID ve CR & T algoritmaları kullanılmış ve ürün verimini etkileyen ana faktörler tanımlanmıştır. Buna dayanarak, çiftçilerin hem hasat mevsimini hem de ürün verimini tahmin etmeleri için bir karar destek sistemi geliştirilmiştir.

*Sorumlu yazar: canan@cs.deu.edu.tr

1. Introduction

Today, studies in the field of agriculture have gained higher significance. Rapid growth that is prevalent in today's world population is one of the reasons for this situation. Agricultural expansion and development in the past forty years have increased the quantity of food produced and improved the quality of fresh food available worldwide [1]. Despite this ongoing expansion, inefficient agricultural practices persist in many parts of the world, because of a lack of modern tools and technologies [2,3]. Agricultural resources should be used properly to prevent the deprivation that might be caused by the rapid population growth and uneven consumption. [4].

Agriculture is an indispensable sector all over the world because it plays a major role in the contribution to the national income and employment, direct and indirect influence on export, raw materials and capital supply to other sectors, and the survival of the country's population. The agriculture sector has been out of the interest of the informatics sector although it has undertaken very important tasks in the economic and social development of the countries. Evidently, technological advancements in agriculture will lead to improvements in productivity.

To analyze the agricultural data, interdisciplinary knowledge that includes math, statistics, crop agronomy, computer hardware and software is needed. The agricultural raw data is transformed by researchers into useful information through data mining. Data mining is the process of discovering previously unknown and potentially valuable patterns in large datasets [5]. Following the conversion of the data into useful information by making use of data mining techniques, computer-assisted decision support systems are developed. This allows farmers to make right decisions and predictions in advance in agriculture. The decision making process for farmers proves that crop yield is directly affected by the climatic characteristics of the region, use of water resources, proper and timely use of pesticides and fertilizers [6].

In this paper, an attempt is made to show how the integration of agricultural data that includes barley yield and climate conditions can be useful for the optimization of harvesting time with data mining. Data from İzmir-Menemen Provincial Directorate of Agriculture and [7] the Turkish State Meteorological Service [8] were used. The data retrieved on an annual basis relates to barley product, and encompasses various features pertaining to the product. The CHAID and the C&R TREE data mining algorithms were applied on SPSS Clementine software in accordance with the data used in the study. Furthermore, a decision support system was developed for farmers to track crop yield and harvest season in advance. The outcomes of the model should have many benefits for optimum agricultural harvesting time management and farming practices in the future.

The remainder of this paper is as follows. Section 2 presents a brief review of related work. Chapter 3 presents the data in general terms and the data mining studies conducted on the data. Chapter 4 explains the Decision Support System developed. A discussion of our results and conclusions is presented in Section 5.

2. Literature Review

Taechatanasat & Armstrong points out that special data is required for creating an agricultural decision support system [6]. This study presents an overview of the agricultural data required for decision making in agriculture. In addition, several important systems are designed for collecting data, and they provide ideas for development. Akın, Yıldırım & Çakan [9] presented that why the decision support systems are not used by poor farmers, contrary to large-scale farms. Sustainability in agricultural production and animal husbandry also emphasized. The goal of the project is to design a system that will boost the efficiency of poor farmers by predicting the risks, and to develop an easy-to-use computer system to this end. Zhu, Zhang, & Sun [10] points out that progress in agriculture will only be through sensitive farming practices. They developed the GIS-based agriculture expert system. This study also covers the data mining and Web technologies that will apply to the agricultural expert decision system. Phoksawat & Mahmuddin [11] developed an optimization model for the maximum profit and to minimize the cost for cultivating the plants. They developed a decision support system that includes the ontology-based knowledge and multi-objective optimization model. Raghuveer, Yogesh, & Shwetha stated that it is quite significant to predict crop yield in advance for market dynamics. They present some of the techniques such as k-means, the k-nearest neighbor, and decision tree in the field of agriculture [12]. Ramesh & Vardhan pointed out that the problem of yield forecast is a major problem that remains unsolved [13]. Different data mining techniques are employed to evaluate the agriculture for estimating the future years' crop production. Their study presents a brief analysis of crop yield forecast using "Multiple Linear Regression (MLR)" technique and "Density-Based Clustering" technique for the selected region in East Godavari district of Andhra Pradesh in India.

3- Algorithms and application

The purpose of this paper is to create a decision support system that will be of help for farmers in their agricultural practices by means of defining the factors that affect crop yield. The ultimate aim is the modeling of various classification algorithms on the data and the observation of the results. This paper intends to design and verify a classification model through the usage of decision trees. This section explains data description, data pre-processing, and the classification model creation using SPSS Clementine software decision trees with CHAID and CR&T algorithms. Together with the results obtained from this analysis, windows forms applications in ASP.NET was designed on Visual Studio 2013. For this decision support system, Microsoft SQL Server 2012 was used as the database to record the information about farmers and the application. Finally, a decision support system was developed for farmers to track crop yield and harvest season in advance.

3.1 Data Description and Data Pre-processing

Original data was obtained from the İzmir Menemen Provincial Directorate of Agriculture regarding Barley product's characteristics. The weather data retrieved from the Meteorological Service Department. In this first step we performed the necessary data cleaning, standardization, and correlation. Data

cleaning techniques allow us to fill in missing values, smooth noisy data, identify outliers, and correct inconsistencies in the data.

The barley product data covers the amounts of barley crop yields of 1077 records over a 10-year period in Menemen, Nazilli, Karacabey and its content is as follows: Variety of barley product, planting date, harvest date, yield, weight in hectoliter, weight by 1000 grams, protein ratio, sieve fraction, earing span, length, location. The weather data of a 10-year period presented in the form of monthly totals covers the following content: monthly total of solar heat, monthly average of humidity, monthly average of wind speed, monthly average of temperature, and monthly total of rainfall. These two types of data were combined and formed as a new set to record all the data. In other words, we generated a data matrix O with the dimensions of 1077x20 with the attributes.

3.2 Decision Tree Algorithms

Decision trees were first proposed by Bierman et. al. in 1984 [14]. This technique is based on when there is an independent variable with the strongest association, the dataset is based on the logic of dividing it into two. In this way, a tree structure is created by continuing until the divisions are completed. A decision tree is a structure that is used by dividing large quantities of records into very small groups of records by applying simple decision-making steps. With each successful partitioning process, the members of the result groups become much more similar to one another. It consists of a root, with any number of nodes, branches, and leafs (also known as terminals). In many classification problems where large databases are used and in decision trees containing complex or inaccurate information, decision trees are a useful solution. Important algorithms such as C4.5, CR&T, ID3, and QUEST, are used for the building of decision trees. In our study, it is observed that classification algorithm is more suitable considering the structure of the data. Thus, in the study, the most suitable options i.e. CHAID Algorithm and C&R TREE Algorithm were respectively used.

3.2.1 CHAID Algorithm

CHAID method [15] uses the chi-square test to generate the optimal partitioning by creating the trees in which each node defines the partitioned state. The CHAID method is easy to interpret and can be used for classification and detection of interaction between variables.

In CHAID algorithm; target value is selected as yield, input values are selected as the weather data, and various rules are then formed accordingly. These rules are subsequently interpreted to yield results from them. The rules shown in Figure 1 below were used to interpret this process.

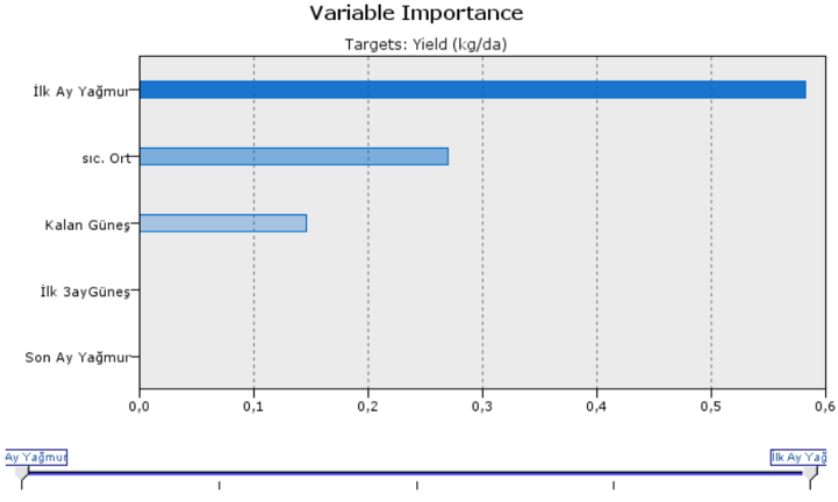
This process reveals the factors that have the highest effect on crop yield based on this algorithm. It is observed that the most important of these factors is the total rainfall for the 1-month period following the crop planting. The factors affecting crop yield are given in order of importance in Figure 2.

If applied again with partition taken based on the location, CHAID algorithm allows different rules to form; and these rules can be interpreted to get new results. The rules shown in Figure 3 below were used to interpret this process.

İlk Ay Yağmur <= 6 [Ave: 507,909, Effect: 78,745] ⇨ 507,909
İlk Ay Yağmur > 6 and İlk Ay Yağmur <= 20 [Ave: 417,628, Effect: -11,536]
  İlk 3ayGüneş <= 235 [Ave: 489,167, Effect: 71,539] ⇨ 489,167
  İlk 3ayGüneş > 235 and İlk 3ayGüneş <= 481 [Ave: 244,8, Effect: -172,828] ⇨ 244,8
  İlk 3ayGüneş > 481 [Ave: 417,846, Effect: 0,218] ⇨ 417,846
İlk Ay Yağmur > 20 and İlk Ay Yağmur <= 31 [Ave: 510,951, Effect: 81,787] ⇨ 510,951
İlk Ay Yağmur > 31 and İlk Ay Yağmur <= 39 [Ave: 344,162, Effect: -85,002]
  sıc. Ort <= 12 [Ave: 328,029, Effect: -16,134] ⇨ 328,029
  sıc. Ort > 12 and sıc. Ort <= 14 [Ave: 443, Effect: 98,838] ⇨ 443,0
  sıc. Ort > 14 [Ave: 346,235, Effect: 2,073] ⇨ 346,235
İlk Ay Yağmur > 39 and İlk Ay Yağmur <= 56 [Ave: 430,358, Effect: 1,194]
  İlk 3ayGüneş <= 386 [Ave: 405,638, Effect: -24,72] ⇨ 405,638
  İlk 3ayGüneş > 386 [Ave: 488,45, Effect: 58,092] ⇨ 488,45
İlk Ay Yağmur > 56 and İlk Ay Yağmur <= 85 [Ave: 641,5, Effect: 212,336] ⇨ 641,5
İlk Ay Yağmur > 85 and İlk Ay Yağmur <= 98 [Ave: 423,029, Effect: -6,135]
  İlk 3ayGüneş <= 352 [Ave: 372,214, Effect: -50,815] ⇨ 372,214
  İlk 3ayGüneş > 352 [Ave: 458,6, Effect: 35,571] ⇨ 458,6
İlk Ay Yağmur > 98 [Ave: 306,659, Effect: -122,505]
  Son Ay Yağmur <= 29 [Ave: 282,375, Effect: -24,284] ⇨ 282,375
  Son Ay Yağmur > 29 [Ave: 335,8, Effect: 29,141]
    Kalan Güneş <= 995 [Ave: 406,5, Effect: 70,7] ⇨ 406,5
    Kalan Güneş > 995 [Ave: 318,125, Effect: -17,675] ⇨ 318,125

İlk Ay Yağmur=Total rainfall in the first month
İlk 3Ay Güneş=Total holar heat in the first 3 months
Sıc. Ort=Temperature average
Son Ay Yağmur=Total rainfall in the last month

Figure 1. The rules by using CHAID algorithm



İlk Ay Yağmur=Total rainfall in the first month
İlk 3Ay Güneş=Total holar heat in the first 3 months
Sıc. Ort=Temperature average
Son Ay Yağmur=Total rainfall in the last month

Figure 2. Variable Importance by using CHAID algorithm

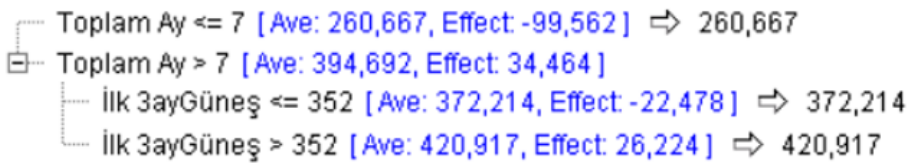


Figure 3. The rules by using CHAID algorithm on the second run

If partition is changed to location-based, the most important factor is observed to be the period that spans from planting date to harvest date. The factors affecting crop yield within this approach are given in order of importance in Figure 4.

Based on these two separate approaches employed in CHAID algorithm, the major factors affecting crop yield are as follows in order of importance: Total months from crop planting to harvesting, total amount of solar heat that the crop gets within the first three months after being planted, and total amount of rainfall within the first months following crop planting.

3.2.2 C&R Tree Algorithm

C&R TREE Algorithm is based on the principle of separating two trees from each decision node. In the C&R TREE algorithm, partitioning is performed by applying a certain criterion in a node. This takes into account the values for which all the qualities are present, and after all matches two divisions are obtained. Selection is performed on these divisions.

In C&R TREE algorithm; target value is selected as yield, input values are selected as the weather data, partition is done on the basis of location, and various rules are then formed accordingly. These rules are subsequently interpreted to yield results from them. The rules shown in Figure 5 below were used to interpret this process.

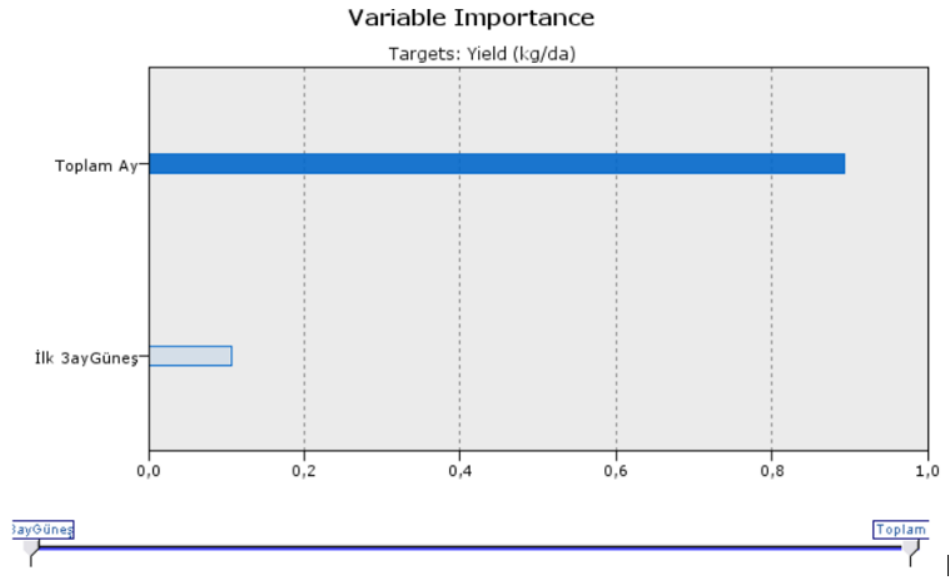
According to C&R Tree Algorithm, the two most important factors affecting crop yield are the total solar heat that the crop gets within the first three months following crop planting, and the amount of solar heat that it gets starting this 3-month period until harvest. It is also observed that these two factors are proportional to each other. Based on this approach, if the ratio of total solar heat that the crop gets within the first three months following crop planting to total solar heat that it gets in the period following this 3-month period does not double it, then crop yield deteriorates. If it doubles, then the crop yield is boosted.

3.3 Results

The CHAID Algorithm and the C&R TREE Algorithm analyses prove that the factors affecting crop yield are as follows in order of importance:

1. Total amount of solar heat that the crop gets within the first 3 months after being planted.
2. Total amount of rainfall within the first months following crop planting.
3. Ratio of the total solar heat that the crop gets within the first three months following crop planting to the total solar heat that it gets in the period following this 3-month period.
4. Total months from crop planting until harvest time.

To help users to make decisions, after obtaining necessary information from the Meteorological Service Department, a decision support system has been created pursuant to these factors.



Toplam Ay=Total month
İlk 3Ay Güneş=Total holar heat in the first 3 months

Figure 4. Variable Importance by using CHAID algorithm on the second run



Kalan güneş=sum of amount of solar heat received until harvest time – sum of solar heat received for the first 3 months
İlk 3Ay Güneş=Total holar heat in the first 3 months

Figure 5. The rules by using C&R TREE algorithm

4. Development of a Prototype Decision Support System

Decision support systems are computer based information systems that help people make decisions. They enable efficient use of data and models, enabling complex problems to be solved more easily. People make mistakes when making decisions on their own.

In this study, to help farmers get maximum amount of crop yield in future, a decision support system has been developed based on farmers' past product information, as well as on meteorological data. The decision support system was developed in Visual Studio using C# programming language and Asp.net software architectural.

The decision support system developed can be used for Menemen, İzmir since the majority of the data used relates to this region. Meanwhile; as the recent weather forecast data made available by the Meteorological Service Department covers 2017-2018, the system can only be used for this period. Implementation helps farmers in three main areas.

4.1 Harvest time forecast

The crop planting date info provided by users and the weather data from the

Meteorological Service Department are combined and used to estimate the harvest time that will bring the highest crop yield. farmers can view the forecasted harvest time based on the planting date provided by the farmers. Currently, this project runs only for the barley yield in İzmir's Menemen district. Farmers can get the information regarding the total amount of solar heat that their crop has received so far, harvest time and the remaining amount of solar heat. This information is created based on the meteorological data available and the results from the data mining, as well as the crop planting date info provided by the farmers themselves. A warning screen is displayed as shown in Figure 6 if the optimal harvest time is missed.

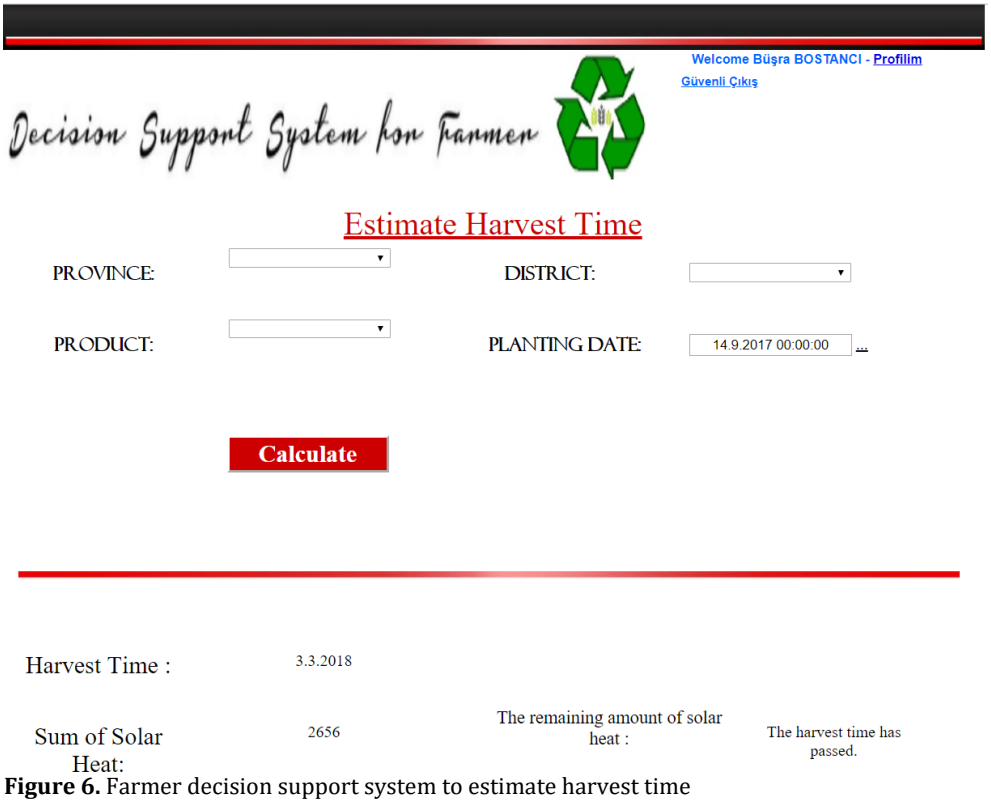


Figure 6. Farmer decision support system to estimate harvest time

4.2 Yield Forecast

The crop planting date and harvest date info provided by users and the weather data from the Meteorological Service Department are combined and used to forecast crop yield amount. On this page, farmers can view the results created based on their planting date and harvest time. Currently, this project runs for the barley product in İzmir's Menemen district. Farmers are informed about the approximate crop yield that they might get after harvesting their crop. This approximate crop yield info is created based on the planting date info and harvest time info provided by the farmers themselves. This page is as shown in Figure 7.

4.3 Farmers' Problems

The problems that farmers face and the questions directed by them to get the information they want are answered by experts. On this page, farmers can ask their questions and access necessary information that they need to contact

page manager. This page is as shown in Figure 8.

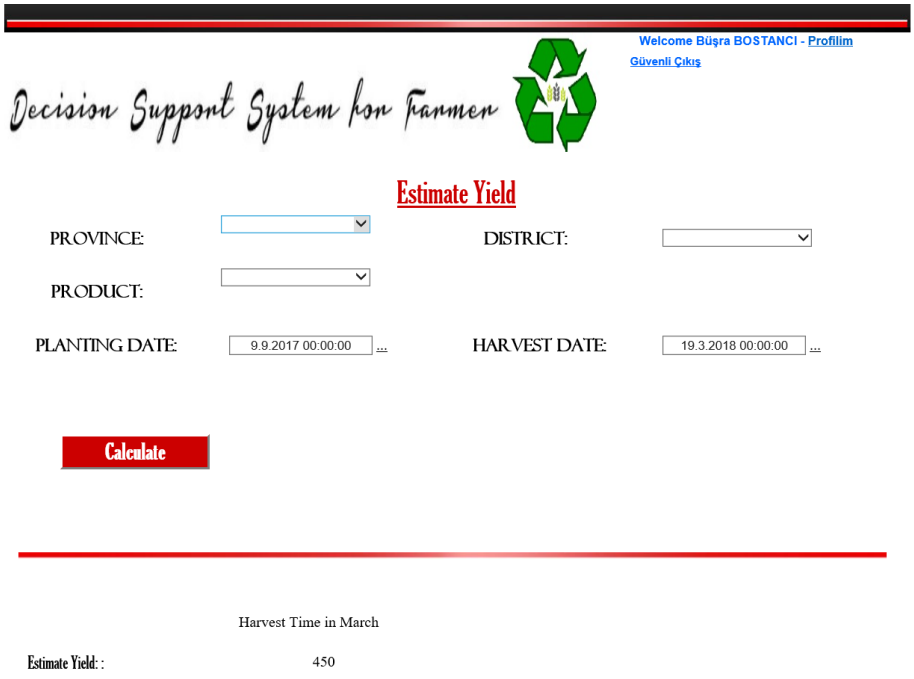


Figure 7. Farmer decision support system to estimate yield

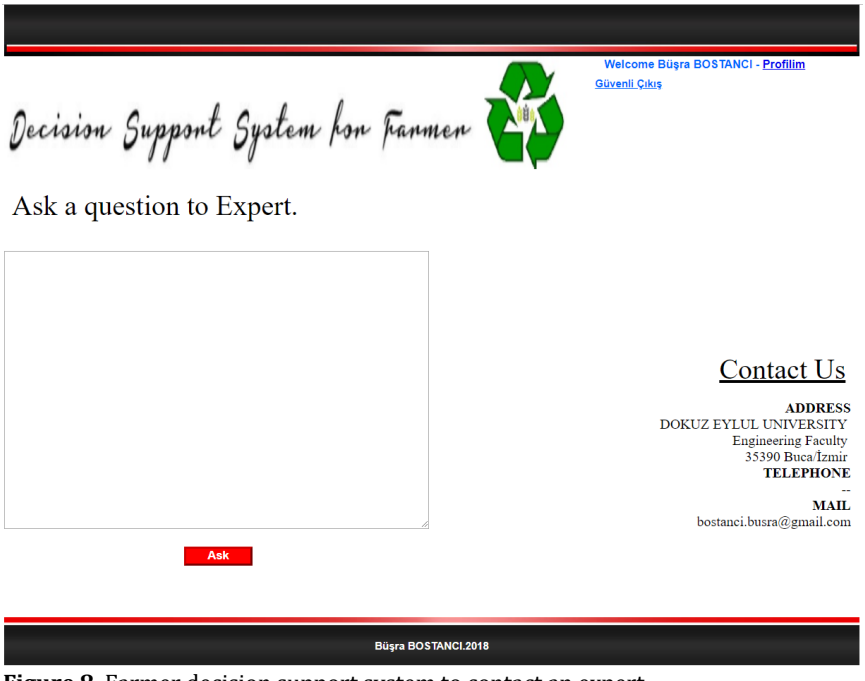


Figure 8. Farmer decision support system to contact an expert

5. Discussion and Conclusion

There is a huge volume of data that is ready to be converted into useful information for the field of agriculture. Farmers need higher amounts of crop yield, lower costs, and minimal damage so they can maximize their gaining which can be achieved through more conscious use of this data at the time of decision-making. In this study, we have applied some sound techniques of data mining to the field of agriculture, which has allowed us to evaluate and examine data for the farmers. Within the scope of this study, the data received from the İzmir Menemen Provincial Directorate of Agriculture was organized for analysis. Then the SPSS Clementine software was used in order to analyze the data with the help of CHAID and CR&T decision tree algorithms to demonstrate the optimum harvesting time for barley yield. This analysis was carried out to determine the factors that have an impact on crop yield. To help farmers make better decisions, an application for a decision support system has been designed in accordance with the results retrieved.

Some meaningful results were observed when the optimum harvesting time and corresponding yield rates were selected as the input variable and as the target variable, respectively. Total amount of solar heat and rainfall that the crop gets within the first three months and the first month following crop planting, respectively, affects crop yield the most.

We have shown how data mining can be successfully applied for this purpose. This work clearly shows that the factors that can be controlled by the farmer can be modeled quite well into a data mining problem such that we can obtains answers to the most common questions, by revealing patterns of interest in otherwise unorganized data.

Acknowledgment

The authors thank the Izmir Menemen Provincial Directorate of Agriculture for sharing data.

References

[1] De Geronimo E, Aparicio VC, Barbaro S, Portocarrero R., Jaime S, & Costa JL (2014). Presence of pesticides in surface water from four sub-basins in Argentina. Chemosphere 107, 423-431.

[2] Laurance WF, Sayer J, Cassman K. (2014). Agricultural expansion and its impacts on tropical nature. Trends in Ecology & Evolution 29, 107-116.

[3] Masters WA, Djurfeldt AA, De Haan C, Hazell P, Jayne T, Jirström M et al. (2013). Urbanization and farm size in Asia and Africa: Implications for food security and agricultural research. Global Food Security 2, 156-165.

[4] Santhosh K.Seelan, Soizik Lagute, Grant M. Cassady, "Remote sensing applications for precision agriculture: A learning community approach," Remote Sensing of Environment, vol. 88, pp. 157-169, 2003.

[5] Fayyad U, Piatesky–Shapiro G, Smyth P. Data mining to knowledge discovery in databases. AI Magazine, 1996. pp. 50-67.

[6] Taechatanasat, P., Armstrong, L. 2013. Decision Support System Data for Farmer Decision Making, Edith Cowan University Research Online ECU Publications.

[7] Gobbett, D., & Bramley, R. (2014). Software tools for precision agriculture Retrieved June 4th, 2014.

[8] Andrew, M., Grundy, M., & Harris, C. (2013). Decision support tools for agriculture.

[9] Tarım ve Hayvancılıkta Bilişim Tabanlı Karar Destek Sistemleri,

Tülin Akın, Coşkun Yıldırım, Handan Çakan.
[10] Research on GIS-based Agriculture Expert System, Zhiqing Zhu, Rongmei Zhang, Jieli Sun, 2009, World Congress on Software Engineering IEEE 252-255.
[11] Ontology-Based Knowledge and Optimization Model for Decision Support System to Intercropping, Kornkanok Phoksawat, Massudi Mahmuddin, 2016 IEEE.
[12] Raghuveer K, Yogesh M J, Shwetha S. Data mining in agriculture: a review AEIJMR 2014; 2: 2348 – 6724.
[13] Ramesh D, Vardhan B V. Analysis of Crop Yield Prediction Using Data Mining Techniques. International Journal of Research in Engineering and Technology 2015; 4: 470-473.
[14] Breiman L., Friedman J., Olshen R., and Stone C. Classification and Regression Trees. Wadsworth Int. Group, 1984.
[15] SPSS Inc*. Clementine® 7.0 User's Guide*, 2002.