Ekonomi, Politika & Finans Araştırmaları Dergisi, 2022, 7(Özel Sayı): 1-18
Journal of Research in Economics, Politics & Finance, 2022, 7(Special Issue): 1-18
Araştırma Makalesi / Research Article, DOI: 10.30784/epfad.1148948

CROP YIELD PREDICTION BY INTEGRATING METEOROLOGICAL AND PESTICIDES USE DATA WITH MACHINE LEARNING METHODS: AN APPLICATION FOR MAJOR CROPS IN TURKEY[*]

Meteoroloji ve Tarım İlacı Kullanım Verilerinin Makine Öğrenmesi Yöntemlerine Entegre Edilmesi Yoluyla Tarımsal Üretim Tahmini: Türkiye'deki Başlıca Mahsuller İçin Bir Uygulama

Hasan Arda BURHAN[**]

Keywords:
Crop Yield Prediction, Machine Learning, Decision Tree Regression, Random Forest Regression.

JEL Codes:
Q16, C15, C5.

Abstract

Agriculture, as one of the most important and vital human activity, is highly vulnerable to global, local and environmental issues. This fragility also surfaced in the initial stages of the COVID-19 pandemic. Accordingly, such matters are considered to have dramatic impacts on demand and pricing dynamics of agricultural products. Nonetheless, improving crop yield and its estimation is the fundamental goal of agricultural activities. To cope with the rapidly changing circumstances, Turkey needs to keep developing data-based agricultural information systems which is also stated as one of the main objectives of the 11th development plan. Therefore, accurate crop yield prediction appears to be a critical task. In this context, using meteorological parameters, pesticides use and crop yield values during 1990-2019, evaluation of machine learning regression methods in the yield prediction of nine major crops in Turkey can be stated as the main aim of this research. After the training, all models are used to predict crop yields and acquired values were compared with actual figures. The results showed that successful predictions were obtained by using the Decision Tree Regression (DTR) and Random Forest Regression (RFR) especially for wheat, barley and maize yields; however, Support Vector Regression (SVR) showed inconsistent predictions.

Anahtar Kelimeler:
Tarımsal Üretim Tahmini, Makine Öğrenmesi, Karar Ağacı Regresyon, Rastgele Orman Regresyon.

JEL Kodları:
Q16, C15, C5.

Öz

En önemli ve hayati insan faaliyetlerden biri olarak tarım, küresel, yerel ve çevresel sorunlara karşı oldukça savunmasızdır. Bu kırılganlık COVID-19 pandemisinin ilk aşamalarında da görülmüştür. Bu bağlamda, söz konusu durumların tarımsal ürünlerin talep ve fiyatlama dinamikleri üzerinde önemli etkilerinin olduğu söylenebilmektedir. Yine de tarımsal faaliyetlerin temel amacı, mahsul verimi ve üretimini iyileştirmek olduğu ifade edilebilir. Türkiye'nin hızla değişen koşullarla başa çıkabilmesi için, 11. Kalkınma Planının da ana hedeflerinden biri olarak belirtilen veriye dayalı tarımsal bilgi sistemlerini geliştirmeye devam etmesi gerekmektedir. Dolayısıyla doğru üretim miktarı tahmini, kritik bir görev olarak öne çıkmaktadır. Bu doğrultuda, 1990-2019 dönemi için meteorolojik parametreler, tarım ilacı kullanımı ve rekolteye dayalı veri setlerini kullanarak, Türkiye'deki dokuz ana mahsulün üretim miktarı tahmininde makine öğrenmesi yöntemlerinin geçerliliğinin değerlendirilmesi, bu çalışmanın temel amacı olarak ifade edilebilir. Eğitim aşamasından sonra tüm modellerle üretim miktarı tahmini yapılmış, elde edilen sonuçlar gerçek değerlerle karşılaştırılmıştır. Sonuçlara göre Karar Ağacı Regresyon (KAR) ve Rastgele Orman Regresyon (ROR) yöntemleriyle, bilhassa buğday, arpa ve mısır için başarılı tahminler alınmış, Destek Vektör Regresyon (DVR) yönteminin ise tutarsız tahminler verdiği görülmüştür.

[*] This study is the revised and edited version of the paper presented at the 6th International Conference on Economics Research and Financial Markets Congress (IERFM)

[**] Asst. Prof. Dr., Kütahya Dumlupınar University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Turkey, arda.burhan@dpu.edu.tr, ORCID: 0000-0003-4043-2652

Makale Geliş Tarihi (Received Date): 26.07.2022 Makale Kabul Tarihi (Accepted Date): 25.09.2022

1. Introduction

Agriculture is one of the most important sources that meets the needs of humanity and is strongly influenced by various factors. Since last two centuries, the significance of agriculture has been boosted by advances in science and technology, changes in the environment and climate, rapid growth in world population and the competitive corporate approach to this activity which is one of the most longstanding way of sustainable food supply (Tauger, 2011: 180). In addition, given the COVID-19 pandemic and its effect upon the global economy, it is also stated that the previously positive medium-outlook for agricultural supply and demand turned into the opposite direction, forcing governments to take effective measures to accommodate themselves to changing circumstances (OECD/FAO, 2020: 63). In this respect, practices of digital agriculture (DA) solutions (i.e. Agriculture 4.0) has gained rapid popularity and utilized as a way to keep up with global trends. As a salient feature of DA, methods and approaches that include collection, transmission and processing of data, applying computer technologies to improve agricultural activity and productivity have started to be used (Tang et al., 2002: 3026). Agriculture also plays an important role in Turkish economy. For instance, in the current 11th development plan for 2019-2023, agriculture is listed as one of the priority development areas. In this plan, the main emphasis is on taking supply-demand balances into consideration in order to create a sustainable, productive and internationally competitive agricultural sector (PSB, 2019: 95). For this reason, data oriented computer technologies are also aimed to be developed in Turkey for efficient agricultural policies (Ozdogan et al., 2017: 185; Ercan et al., 2019: 262-264; Kilavuz and Erdem, 2019: 136; Kirmikil and Ertas 2020: 9-10).

A convenient approach to identify rules, relationships and patterns in a dataset, machine learning (ML) has been applied in many areas such as credit analysis, image recognition, meteorology, medicine, fraud detection, customer relations, bioinformatics as well as in agriculture. (McQueen et al., 1995: 275; Liakos et al., 2018: 2674; Patrício and Rieder, 2018: 69). In ML applications, datasets are generally split in two parts such as a training set and a test set. Additionally, the model itself can be predictive for forecast purposes and/or descriptive to gain knowledge from the utilized data (Klompenburg et al., 2020). In this context, the training set is used to optimize the model which can be defined up to several parameters, while test set is reserved to evaluate the model performance. Thereby, as a component of the artificial intelligence (AI) ML models gain the ability to learn and adapt to the changing dynamics in the examined systems by constructing proper and handy approximations (Alpaydın, 2004: 2). With a remarkable prediction potential, ML approaches deal successfully with agricultural structures which can be both linear or non-linear by nature (Rashid et al., 2021: 63408). As a data focused methodology, ML has yielded significant opportunities in the agricultural domain and there is a vast literature on ML applications in agriculture. Furthermore, according to Liakos et al. (2018), more than 60% of the studies that utilize ML applications in the agricultural literature are related to the crop management that includes disease and weed detection, crop quality and crop yield prediction. Among them, crop yield prediction in particular is regarded as one of the most significant and interesting topics in nowadays agriculture which concerns farmers, traders, policy-makers, agronomists etc. to make wise supply-demand decisions in their activities (Paudel et al., 2021). Not only because of the importance of making accurate predictions while making decisions, but also correct determination of the factors that have influence on crop yields turns out to be possible while working on these forecasts (Vanli et al., 2020: 1757; Pant et al., 2021: 10922). In this respect, factors or parameters such as environmental conditions, climate, weather, crop genotypes, soil, managerial practices, variety of seeds, fertilizer and pesticides use, etc. are mentioned as main determinants of crop yield predictions by considering the related studies in the literature (Benos et al., 2021: 3763; Pant et al., 2021: 10922).

It is stated that there is not a standard dataset for agricultural research in general as it depends on the region, crop types and various factors that are given above (Gopal and Bhargavi, 2019b). Considering these factors, features of weather conditions such as precipitation, temperature, drought, etc. are basic concerns for all actors in agriculture (Jeong et al., 2016: 1). Particularly, variations in climate have been defined as a crucial determinant that has a negative effect on certainty of predictions of crop yields (Lobell and Burke, 2008: 2). This changeability is connected with magnitude and patterns of rainfall, decreases and increases in temperature, wind power and such (Shook et al., 2021). Therefore, taking these factors into consideration and integrating variables of weather conditions in crop yield assessments will lead to timely and accurate predictions in this respect. However, each crop has its own growing process and the importance of each meteorological factor is different. For this reason, when adding these determinants into the prediction models, an integrated approach is recommended to be followed by including more than one of these weather condition variables (Xu et al., 2019: 944). Accordingly, numerous studies have focused on the effects of meteorological parameters on crop yield, such as Lobell and Burke (2008), Jeong et al. (2016), Trnka et al. (2016), Xu et al. (2019), Kang et al. (2020), Pant et al. (2021), Shook et al. (2021) and Zarei et al. (2021). In addition, use of pesticides can be stated as another factor that has influence on crop growth (Pant et al., 2021: 10923). Agricultural production levels are increased in many countries by using modern cultivation methods and soil-water management techniques, but this positive outlook has a reverse side, which is open vulnerability to pests (Oerke, 2006: 39). Although there are ongoing debates about the necessity of pesticides use by considering the risks for the human health and environment, general consensus is that it provides a certain protection and maximizes crop yield (Washuck et al., 2022: 1765). Therefore, it is clear that, as an effective agronomic input and a precaution method, conscious use of pesticides decrease requirements for land use and boost productivity (OECD/FAO, 2020: 38). There are also various studies in the agriculture literature that focuses on the relationship between pesticides use and crop yields (e.g. Toscano et al., 1982; Alston et al. 1993; Oerke, 2006; Kawasaki and Lichtenberg, 2015; Lamichhane, 2017; Xie et al., 2019; Washuck et al., 2022).

The main aim of this study is to apply some of the well-known ML regression methods to predict crop yields in order to evaluate these methods' usability in similar forecasts and offer an alternative approach for future researches. Therefore, by taking the related literature into account, a dataset including crop yields in addition to certain meteorological values and pesticides use figures is constructed to make predictions by using the Support Vector Regression (SVR), Decision Tree Regression (DTR) and Random Forest Regression (RFR) ML methods. The goal of this study is also related with Turkey's data-based and digitized agricultural information systems target which is stated on the current 11th development plan for 2019-2023. Thereby, main motivations of this study can be stated as follows: Covering both meteorological and pesticide usage influences in crop yield predictions since there is a very limited number of studies in the literature that focus on these factors altogether. Secondly, by considering the above mentioned development plan goal, taking a step in filling the gap in Turkish agricultural literature by using ML prediction methods, which have not been commonly used up to now. With regard to the proposed model, the dependent variable set consisted of nine major and most produced crops in

Turkey as apples, barley, grapes, maize, olives, potatoes, sugar beet, tomatoes and wheat which are determined by considering the crop production reports of Turkish Statistical Institute. In accordance with the related studies, meteorological factors were taken as average rain, average temperature, minimum temperature, average wind speed, in addition to the pesticides use data which altogether formed the independent variables set. By using the data for the period 1990-2019, predictions were made for all crops for 2019 by using above mentioned ML approaches and compared with actual observed values in order to evaluate the accuracy of these methods.

Accordingly, the contributions of this study to the existing literature can be stated as follows: Firstly, it proposes a prediction model that includes two very important factors together that have significant influence on crop yields. In addition, three of the well-known ML regression methods have been utilized and evaluated in this study, which are rarely used in crop yield predictions when Turkish agriculture is considered. Also, offering an alternative approach to the current models in crop yield predictions that can be employed by policy-makers, academics, researchers and officials' that work in the field of agriculture regardless of the geographical location and agricultural structure of their country can be specified as another contribution of this study. Additionally, obtained results can be analyzed in detail and strategies can be set considering each crop's different biological and environmental needs in particular. Therefore, based on the model that is proposed by this study, more advanced prediction models can be built by decision-makers. Moreover, government agencies and chambers that support local farmers can develop additional and novel policies by using this yield prediction model which can be regarded as an alternative to the commonly used methods such as, satellite image-based calculations, linear regression analysis, etc. In this context, a reasonable contribution can be provided to the modernization and digitalization of agricultural systems all around the world. The remainder of the paper is organized as follows: a review of the related literature is given in Section 2; Section 3 provides brief explanations of the data and methods; results are presented in Section 4; Section 5 includes discussion and conclusions.

2. Literature Review

Agriculture has been one of the most significantly transformed areas in most countries over the years. Along with the growth in population and increasing demand for agricultural products, states felt the pressure to modernize their farming methods by integrating innovative digital approaches and technologies -which is also called DA or Agriculture 4.0- in order to expand the productions, increase efficiency and maintain sustainability (Araújo et al., 2021: 668). In this context, predicting crop yield presents a challenging task due to its complex structure which includes various factors that affect these forecasts (Shook et al., 2021). As far as these predictions are concerned, it can be stated that ML algorithms deliver encouraging results by learning mutual interactions in a dataset (Paudel et al., 2021). Since its appearance as a branch of AI, ML methods have been extensively used in the field of agriculture and crop yield predictions. Accordingly, various reviews of literature have been published recently such as, Chlingaryan et al. (2018), Liakos et al. (2018), Patrício and Rieder (2018), Klompenburg et al. (2020), Benos et al. (2021), Rashid et al. (2021), Bali and Singla (2022). Considering these reviews, most frequently used ML methods can be stated as the Decision Trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Deep Neural Network (DNN), k-Nearest Neighbor (kNN), SVR, RFR, DTR and Gradient Boosting Tree (GBT) (Chlingaryan et al., 2018: 63; Klompenburg et al., 2020;

Benos et al., 2021: 3773). In accordance with the methodology of this study, selected papers that utilized SVR, DT and RF approaches in crop yield prediction are given in Table 1:

Table 1. SVR, DTR and RFR Applications in Crop Yield Prediction

Reference	Title	Method Used
Rahman et al. (2014)	Machine learning facilitated rice prediction in Bangladesh	DT, ANN, Linear Regression
Jeong et al. (2016)	Random forests for global and regional crop yield predictions	RF, Linear Regression
Everingham et al. (2016)	Accurate prediction of sugarcane yield using a random forest algorithm	RF
Gandhi et al. (2016)	Rice crop yield prediction in India using support vector machines	SVR
Ahmad et al. (2018)	Yield Forecasting of Spring Maize Using Remote Sensing and Crop Modeling in Faisalabad-Punjab Pakistan	SVR, RF, DT
Charoen-Ung and Mittrapiyanuruk (2018)	Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning	RF
Khanal et al. (2018)	Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield	RFR, ANN, SVR, GBT, Cubist
Shah et al. (2018)	Smart farming system: crop yield prediction using regression techniques	SVR, RF, Multivariate Polynomial Regression
Filippi et al. (2019)	An approach to forecast grain crop yield using multi-layered, multi-farm data sets and ML	RF
Gopal and Bhargavi (2019a)	Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms	ANN, SVR, kNN, RF
Xu et al. (2019)	Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China	SVR, RF
Khosla et al. (2020)	Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression	ANN, SVR
Leo et al. (2020)	Predicting within-field cotton yields using publicly available datasets and machine learning	RF. GBT
Dang et al. (2021)	Autumn crop yield prediction using data-driven approaches: Support vector machines, random forest, and deep neural network methods	SVR, RF, DNN
Pant et al. (2021)	Analysis of agricultural crop yield prediction using statistical techniques of machine learning	GBT, RFR, SVR, DT
Paudel et al. (2021)	Machine learning for large-scale crop yield forecasting	kNN, GBT, SVR
Lischeid et al. (2022)	Machine learning in crop yield modelling: A powerful tool, but no surrogate for science	RF, SVR
Paudel et al. (2022)	Machine learning for regional crop yield forecasting in Europe	kNN, GBT, SVR

In the second step, related literature is reviewed for Turkish agricultural sector and it can be stated that a very limited number of studies were found for crop yield predictions, let alone studies that include ML applications. Simsek et al. (2007) estimated the wheat yield by using AgroMetShell model which is developed by Food and Agriculture Organization of the United Nations (FAO) in order to evaluate the effect of meteorological conditions on crops. By utilizing soil, phenological observation, crop coefficient, meteorological and Normalized Difference

Vegetation Index (NDVI) data for cities in Turkey, accurate wheat yield forecasts were obtained for 2005 and 2006. Varjovi and Talu (2016) applied an integrated the Gaussian mixture-ANN model to predict apricot yield by using 1170 video images. The results are regarded as successful with a R-squared value of 0,77. Basakin et al. (2020) applied the Wavelet Fuzzy Time Series (WFTS) and the Gray Prediction (GP) algorithms in order to predict the wheat yield in Turkey by using the dataset for period 1941-2018. To assess model performances, mean square error and coefficient of efficiency success criteria are considered. According to the results, the WFTS models revealed accurate predictions. To forecast wheat yield and area estimation, Vanli et al. (2020) used satellite images where the SVM, RF, DT, kNN and boosting algorithms were trained in order to be used for the spatial distribution of wheat. In addition, a Principal Component Analysis (PCA) model is developed for yield forecasting and a LASSO regression for the coefficient. According to the results, accurate estimations were obtained. Bregaglio et al. (2021) presented a yield prediction system for hazelnut called HADES (HAzelnut yielD forEcaSt) that integrates ML techniques and process-based modelling. Ground observation and meteorological data between 2004-2019 are used along with the hazelnut yield figures. After applying the method which also includes a RF approach, it is stated that HADES method has balanced predictive ability which provides robust and timely information. Kaya and Polat (2021) predicted the wheat yield of southeast Şanlıurfa for 2018-2019 by using NDVI, Modified Soil-Adjusted Vegetation Index (MSAVI) and Green Normalized Difference Vegetation Index (GNDVI) that are acquired from yield figures by parcels and satellite images. According to the results, NDVI yielded the best results with an 82% NDVI value.

3. Data and Methodology

3.1. Data Description and Preprocessing

The data used in the analysis includes a set of variables as crop yields, pesticides use and meteorological parameters. The crop yield dataset contained observed annual yield between 1990 and 2019 for nine major crops in Turkey, namely apples, barley, grapes, maize, olives, potatoes, sugar beet, tomatoes and wheat. Data were taken from the FAO database and as an example, Figure 1 shows the yield plot for sugar beet and tomatoes over the related years:

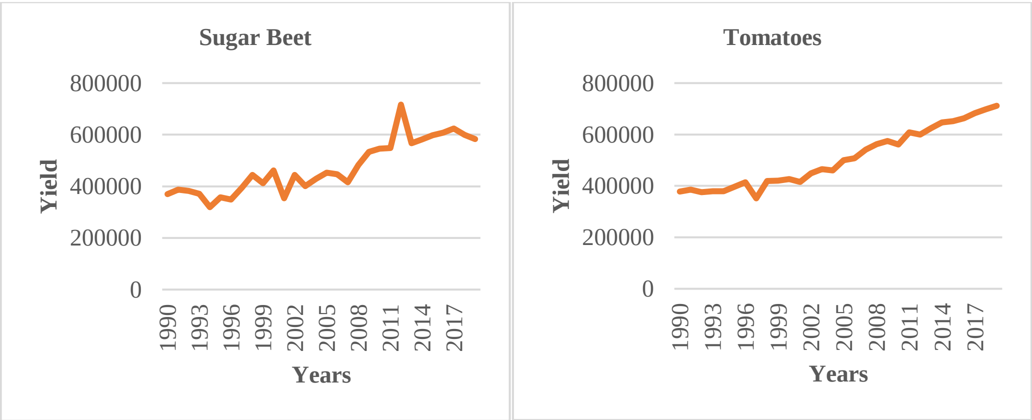

Figure 1. Sugar Beet and Tomatoes Yield Between 1990-2019
Source: FAOSTAT

The pesticides use data included the annual pesticides use in Turkey for the same period of time and the data were acquired again from the FAO database. Figure 2 shows pesticides use in Turkey for the 1990-2019 period:

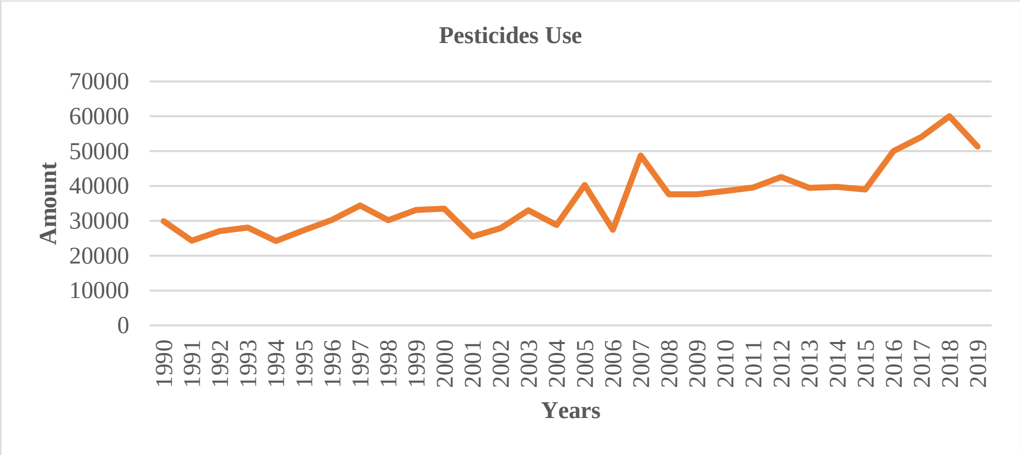

Figure 2. Pesticides Use Between 1990-2019
Source: FAOSTAT

The meteorological data included annual record of four variables, namely average rain, average temperature, minimum temperature and average wind speed. The related data was obtained from Turkish State Meteorological Service reports

The dataset did not have any missing values, therefore only feature scaling of the data for SVR application is performed in the data preprocessing stage. It is stated that compared to the performance of without scaling, the SVR yields better results with a scaling approach (Lin et al., 2018: 123). However, opposed to the SVR, there is no requirement for feature scaling in tree-based models such as the DTR and RFR (Liu et al., 2021: 3).

3.2. Methods

3.2.1. Support Vector Regression (SVR)

In a feature space with high-dimension, the main idea of the SVR is to compute a function of linear regression in which the data are mapped by a non-linear function (Basak et al., 2007: 203). In this context, the main aim of the SVR can be defined as finding a function for all the data used for training which has most deviation from the actual value *y*. Since the square of errors used in regression as;

$$e_2(r^t, f(x^t)) = [r^t - f(x^t)]^2 \quad (1)$$

in SVR, an ε sensitive loss function;

$$e_\varepsilon(r^t, f(x^t)) = \begin{cases} 0 & \text{if } |r^t - f(x^t)| < \varepsilon \\ |r^t - f(x^t)| - \varepsilon & \text{otherwise} \end{cases} \quad (2)$$

meaning that errors beyond having a linear effect and valued up to ε are tolerated (for further details, see Alpaydın, 2004: 226).

3.2.2. Decision Tree Regression (DTR)

By using the training data and constructing a decision tree where each feature is presented by a node, the DTR predicts the target variable and presents an easy interpretation opportunity since it provides the results in a tree structure (Millán-Castillo et al., 2020: 4124). Decision trees can be used both for classification and also for regression. By using a binary split, the algorithm separates the data into two parts to minimize sum of squared deviations from the mean in each part until all of the nodes attain a minimum node size which is specified by the user (Xu et al., 2005: 323). In below algorithm steps, a pseudo code for construction of a DTR is given (Pekel, 2020: 1114):

(i) Starting from a single node,
(ii) For each feature *X*, fitness function value *S* is obtained and a split that provides minimum values of *S* is selected,
(iii) For all new nodes apply step 2 and exit when stopping criterion is met. (see Alpaydın, 2004: 180-185 and Pekel, 2020: 1112-1114 for details)

3.2.3. Random Forest Regression (RFR)

As a method both used for regression and classifications, random forests are mainly ground on the results of various decision trees (Lischeid et al., 2022). These models bootstrap training data to output these trees and generates a mean prediction for a regression or a class for a classification after collecting results from all of them, which is also called *bagging* (Shah et al., 2018: 53). With the help of bagging, subsamples are generated from the original dataset and brings out predictors from each and uses averaging for decisions (Scornet et al., 2015: 1717; Schwalbert, et al., 2020). Because of the random processes added to the steps of the algorithm, decision trees differ from each other. Also a decrease in the variance of the prediction can be seen which improves the performance of the approach (Zuo et al., 2020: 1280). It is also stated that the RFR is robust when overfitting and nonlinearity is considered unlike linear regression (Khanal et al., 2018: 217). Simple steps of the RFR algorithm are given as follows (Chen et al., 2020: 5743):

(i) By using bootstrap sampling, data is obtained from the training set,
(ii) Random selection of *m* features is performed for each node and in accordance with the Gini coefficient, optimal feature is selected,
(iii) Prediction accuracy is determined by comparing outcomes with the test data,
(iv) By considering the estimation error, optimum number of trees in the algorithm is specified and the model should be rebuilt accordingly,
(v) With new datasets, average estimations for all trees are obtained, which is the final output.

4. Results

In this study, three ML regression methods were implemented by using Python programming language and related tools in the Sci-kit library to train the prediction models that include crop yield values in addition to meteorological and pesticides use data. Namely, the SVR, DTR and RFR algorithms were utilized for yield prediction of nine major crops of Turkey. But in the first phase, models were evaluated. Steps of each process is given in Figure 3.

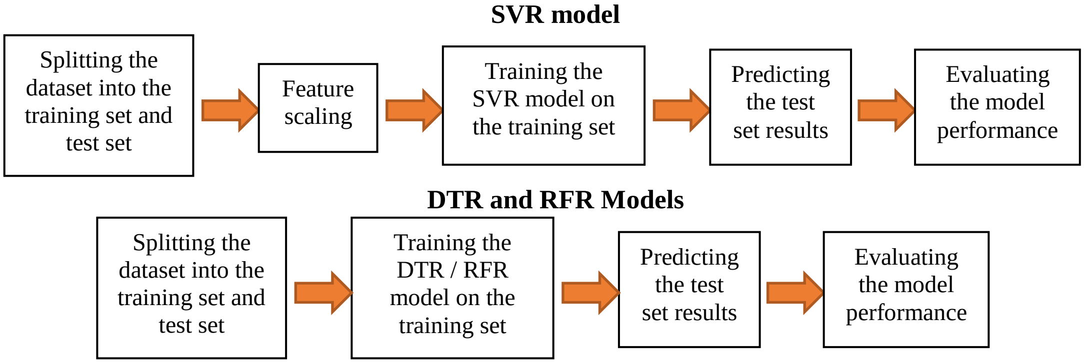

Figure 3. Evaluation Steps of the Models

As stated above in Section 3, feature scaling by using standard scaling was performed only for the SVR model. After this step, all of the models were trained by using the training sets. Partial results of the training applications are given in Table 2:

Table 2. Partial Training Results of the Models

SVR		DTR		RFR	
Prediction	**Actual**	**Prediction**	**Actual**	**Prediction**	**Actual**
74.884	250.409	230.653	250.409	258.611,5	250.409
219.417	426.563	461.098	426.563	439.205,9	426.563
308.097	59.896	65.256	59.896	63.606,6	59.896
65.301	323.303	309.489	323.303	313.895,9	323.303
130.829	42.489	37.322	42.489	38.753,3	42.489
53.555	276.422	328.252	276.422	314.151,1	276.422
57.126	324.475	324.041	324.475	316.426,5	324.475
154.624	446.448	429.047	446.448	411.825,9	446.448
52.712	18.048	17.358	18.048	248.92,9	18.048
84.453	66.038	67.308	66.038	67.689,7	66.038
27.229	45.455	42.427	45.455	42.369,4	45.455
50.454	71.988	68.381	71.988	76.940,5	71.988
403.448	81.163	85.569	81.163	83.377,8	81.163
417.585	91.906	91.593	91.906	93.079,6	91.906
143.205	207.565	168.739	207.565	161.212,5	207.565
172.422	24.294	28.803	24.294	26.101,2	24.294
66.255	30.000	7.755	30.000	11.345,8	30.000
132.459	11.776	9.734	11.776	10.531,6	11.776
87.362	416.115	354.240	416.115	383.591,8	416.115
82.692	22.564	20.765	22.564	21.534,3	22.564

After splitting the dataset and performing feature scaling, training results were acquired as partially given above in Table 2 in which prediction and actual values are compared. According to Table 2, it can be stated that most of the predictions obtained by the SVR method have significant differences compared to the actual values, which are given on the second column next to the predictions. On the other hand, calculations with the DTR approach yielded relatively accurate results. For example, according to the first row of the table, the predicted value is around 230.000 while the actual value is around 250.000. Finally, obtained results with the RFR method are also almost accurate, since compared prediction and actual values are mostly close to each other except some figures in a few rows.

Afterwards to evaluate the models, R-squared values were calculated (for detailed explanations regarding evaluation approaches see Lobell and Burke, 2008: 2; Shah et al., 2018: 54; Pant et al., 2021: 10925) and results are given in Table 3 below:

Table 3. R-squared Values of the Models

Models	**Values**
Support Vector Regression	-0,296
Decision Tree Regression	0,966
Random Forest Regression	0,976

According to Table 3, the RFR and DTR models yielded the highest accuracy with R-squared values of 0,976 and 0,966 respectively, while the SVR showed the poorest performance. In the next phase, all three models were used to predict crop yields. The actual values on the last year of the dataset which is 2019, was taken as validation values. For 2019 predictions, training data included values from 1990 to 2018. Prediction steps for each model is given in Figure 4:

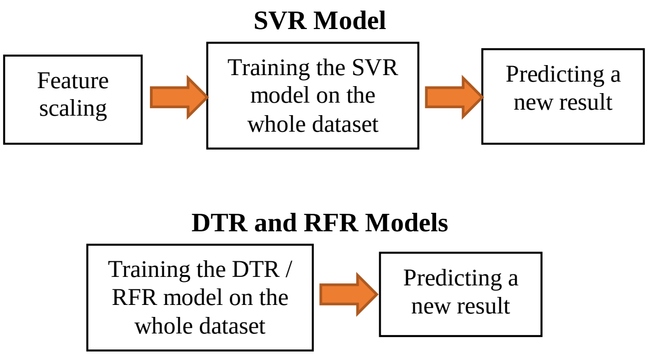

Figure 4. Prediction Steps of the Models

In Table 4, performances of the three models are compared by considering the yield predictions and actual values of the nine crops:

Table 4. Prediction Performance of the Models for 2019

	SVR		**DTR**		**RFR**	
	Prediction	**Actual**	**Prediction**	**Actual**	**Prediction**	**Actual**
Grapes	289.161,27	101.125	94.307	101.125	87.417	101.125
Sugar Beet	277.056,85	583.216	608.559	583.216	624.850	583.216
Tomatoes	242.122,79	711.767	663.088	711.767	663.910	711.767
Wheat	191.972,14	27.811	**27.440**	**27.811**	**26.521**	**27.811**
Apples	139.881,28	207.451	144.702	207.451	149.196	207.451
Barley	100.779,38	26.565	24.815	26.565	**25.002**	**26.565**
Maize	86.376,38	94.034	90.748	94.034	**93.187**	**94.034**
Olives	101.340,13	17.346	21.402	17.346	20.328	17.346
Potatoes	142.007,24	353.766	336.014	353.766	323.385	353.766

Considering the outcome in Table 4, it can be stated that the DTR and RFR models yielded almost successful results. For example, while the actual observed yield value for grapes is 101.125 in 2019, the DTR method predicted a value of 94.307 and around 87.000 with the RFR. For tomatoes, barley and olives, both methods also provided very close amounts to each other. For wheat, a very close prediction was acquired with the DTR approach. Subsequently, good predictions for wheat and barley, and an almost successful results for maize is acquired with the RFR approach. By considering the related literature, successful wheat predictions for Turkey have also been obtained by other ML methods in Simsek et al. (2007), Basakin et al. (2020), Vanli et al. (2020) and Kaya and Polat (2021). However, a comparison of prediction values could not be performed since none of this studies presented predictions for 2019. As stated previously, there are very a limited number of studies that include prediction of crop yields with ML methods in Turkish agricultural literature. When analyzed, it was seen that remaining studies performed forecasts for different products than this study, such as apricot and hazelnuts. With regard to the remaining results, relatively good figures were calculated for potatoes and sugar beet. The least successful results were obtained for apples. In addition, SVR algorithm showed inconsistent predictions for almost all crops as this model also yielded a negative R-squared value in Table 3, which indicates an insignificant prediction and a low representation capability for the related dataset (Vaid and Ghose, 2020: 340; França et al., 2022). As a result, it can be stated that the most accurate results are obtained with DTR and RFR applications especially for wheat, barley and maize yields. When compared to related literature, it can be stated that successful yield predictions with DTR and RFR methods have also been obtained in the studies of Everingham et al. (2016) for sugarcane, Ahmad et al. (2018) for maize, Shah et al. (2018) for corn, Pant et al. (2021) for maize, potatoes, rice and wheat.

According to the actual values included in the entire dataset, Turkey's wheat production was 27.440 in 2018 and this amount increased to 27.811 in 2019. By taking the most accurate predictions into account, it can be stated that none of the successful methods in this study have predicted this increment in 2019 for wheat production, since DTR have predicted that in 2019, the wheat yield value will remain unchanged (as 27 .440), and RFR approach forecasted a decrease in 2019 compared to the previous year. On the other hand, a slight decrease in barley yield from 26.911 in 2018 to 26.565 in 2019 has been noticed in the actual dataset; however, RFR method -which is the most accurate one for barley yield, has predicted a significant decline to 25.002 units. When maize values are considered, observed 2018 value was 96.358, which is followed by a remarkable fall in 2019 to 94.034. But according to Table 4, a much lower yield value, 93.187 is obtained with the RFR method. In this regard, although close predictions are calculated for 2019 with the tree-based models, none of them displayed efficient performance in predicting the progress of wheat, barley and maize yields when last two years of the dataset are compared.

5. Discussion and Conclusions

It is a well-known fact that agriculture is the main source of nutrition and livelihood of humanity. Since the world population is anticipated to reach about 9,7 billion in 2050, agricultural production must be increased between 40 to 54 percent than the last ten years (FAO, 2021: 4). But agricultural systems vary in terms of performance, technologies and applicability across the world. Therefore, it can be clearly stated that it proceeds with intense effort and has a significant vulnerability to various environmental, climatic and operating conditions. For instance, current abnormalities in the weather will be boosted in the future by the negative impacts of ongoing climate change and new measures will be required to construct a sustainable agricultural practice

(Araújo et al., 2021: 667). Therefore, optimization of these practices by delimiting the burden of above mentioned conditions comes forward as a crucial task for the parties in the agriculture industry (Benos et al., 2021: 3758). The digital agricultural revolution, which is also called Agriculture 4.0, is mainly inspired by Industry 4.0 with its associated technological innovation and advances. With the collection of a large amount of data and using of big data analytics through information and communication technologies such as cloud computing, AI, Internet of Things (IoT) etc., several goals are pursued in the agricultural sector as, increasing process and production quality, maintaining sustainability and enhancing crop yield. (Shi et al., 2019; Zambon et al., 2019: 9). When these are considered, one of the major and complex one can be stated as crop yield prediction and its complexity mainly depends on environment, climate, genotype and interactions of these factors (Khaki and Wang, 2019: 1-2).

Over time, labor-intensive and conventional methods in addition to crop growth models, remote sensing methods, surveys and various statistical models are used in crop yield estimations. But recently, ML methods such as DT, RF, ANN provide accurate outcomes in agriculture and crop yield predictions in particular, with its data-driven approach that gain insights from data by learning relationships among its elements (McQueen et al., 1995: 275; Paudel et al., 2022). Since, Turkey set a goal of using digitized and data-based methods for creating a competitive, productive and sustainable agriculture sector in the current development plan, it can be stated that ML methods should be integrated into the existing practices to generate desired outcomes. In this context, a ML approach is presented in this study as an alternative for crop yield predictions by considering meteorological and pesticides use data for the period 1990-2019. The applied methodology included the use of three different ML algorithms as SVR, DTR and RFR for predictions of nine major crops in Turkey. In order to evaluate the accuracy of the models, observed crop yield values of 2019 were used for validation and obtained predictions were compared to the figures belong to that year. Following the steps of each algorithm, results indicate that DTR and RFR are the best performing models and they both outperformed the SVR model for this application. Prediction results for the SVR model were inaccurate and this can be anticipated because of the negative R-squared value which was calculated in the model evaluation phase. Considering the results, it can also be stated that meteorological data combined with pesticides use data provide useful information regarding yield predictions of these crops and these factors and methods can be suggested to be included in similar predictions.

This study has some important limitations regarding the dataset and the applied methodology. First of all, a variety of different crops were included in this study which naturally expected to have their own characteristic needs and interactions when the meteorological parameters and pesticides use are concerned. Additionally, crops have different growth stages and determinants of this stages vary in accordance with specific features of each particular crop. Moreover, although several other climatic factors, such as CO_2 concentration, humidity, etc. may have an impact on crop yields, only temperature, precipitation and wind based variables were included in the models. Therefore, it can be stated that this study did not fully consider above mentioned matters. Another limitation is that the training of the algorithms and predictions were performed on a limited number of datasets and there is no doubt that a larger dataset would have a positive influence on the prediction capability of these models. Finally, only three of the basic ML regression methods were used in this study; however, several other ML methods such as ANN, DNN, kNN are widely and effectively used for crop yield predictions in the recent literature.

The main contributions of this paper can be stated as combining two of the most important factors in crop yield predictions as meteorological variables and pesticides use in the same model. Additionally, an up-to-date methodology is adopted based on three well-known ML regression algorithms which is not common especially in Turkish agricultural literature. Thereby, it can be stated that this study presents an addition of a new approach to Turkey's existing agricultural literature which utilizes meteorological, pesticides use and yield data for the evaluation of some ML regression methods in crop yield predictions since almost successful predictions were acquired with tree-based models (the DTR and RFR) with both training and test sets, and the closeness of predicted and actual numbers shows a good fit.

Regarding the policy implications of this research which evaluates and confirms the applicability of certain ML regression methods in agricultural yield prediction, the proposed prediction model can be utilized both by decision and policy-makers as well as other officials in the agriculture sector in Turkey in accordance with the emphasis on the 11th development plan for developing advanced information systems with digitalized and data-based approaches in agriculture. More specifically, decision-makers and other responsible parties in wheat, barley and maize production can take advantage of the proposed approach in this study, since very close yield predictions were obtained for these agricultural products. Therefore, as wheat, barley and maize are predominantly produced in the in Central and Southeast Anatolia Region in Turkey, government agencies affiliated to Turkish Republic Ministry of Agriculture and Forestry in these regions may mostly benefit from the outcomes of this research. In this context, utilized methods in this study can also be introduced to related authorities all around the world for further applications. Furthermore, the proposed model and methodology can be taken as a basis by all policy-makers and officials, as it can be improved and enhanced by considering each crop's different climatic, pesticide needs and using other various ML methods with much bigger datasets. Thereby, as another policy implication, useful forecasts can be obtained not only for the crops that yielded close predictions in this study, but also for every other agricultural product in the globe. By this means, digital agriculture goals of countries and aims of Turkey's current development plan for 2019-2023 period in specific can also be supported by taking this yield prediction approach and model into consideration and new policies can be developed accordingly.

For future studies, scope and methodology of this analysis can be expanded and modified by using other ML algorithms with larger datasets. Thereby, one of the stated limitations of this study can be overcome. As stated previously, considering each crop's unique biological needs separately, constructing models that include other types of agricultural products, various other chemicals used in agriculture, different meteorological determinants and other variables that have influence on crop yield predictions may help researchers in the future to handle other limitations of this study.

Declaration of Research and Publication Ethics
This study which does not require ethics committee approval and/or legal/specific permission complies with the research and publication ethics.

Researcher's Contribution Rate Statement
I am a single author of this paper. My contribution is 100%.

Declaration of Researcher's Conflict of Interest
There is no potential conflicts of interest in this study.

References

Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Rahman, M.H., Ahmad, A. and Judge, J. (2018). Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. *Journal of the Indian Society of Remote Sensing*, 46, 1701-1711. https://doi.org/10.1007/s12524-018-0825-8

Alpaydın, E. (2004). *Introduction to machine learning*. Cambridge: The MIT Press.

Alston, D.G., Schmitt, D.P., Bradley, J.R. and Coble, H. (1993). Multiple pest interactions in soybean: Effects on heterodera glycines egg populations and crop yield. *Journal of Nematology*, 25(1), 42-49. Retrieved from https://www.ncbi.nlm.nih.gov/

Araújo, S.O., Peres, R.S., Barata, J., Lidon, F. and Ramalho, J.C. (2021). Characterising the agriculture 4.0 landscape—emerging trends, challenges and opportunities. *Agronomy*, 11(4), 667-703. https://doi.org/10.3390/agronomy11040667

Bali, N. and Singla, A. (2022). Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. *Archives of Computational Methods in Engineering*, 95, 95-112. https://doi.org/10.1007/s11831-021-09569-8

Basak, D., Pal, S. and Patranabis, D.C. (2007). Support vector regression. *Neural Information Processing*, 11(10), 203-224. Retrieved from https://citeseerx.ist.psu.edu/

Başakın, E.E., Ekmekcioglu, Ö., Özger, M. and Çelik, A. (2020). Prediction of Turkey wheat yield by wavelet fuzzy time series and gray prediction methods. *Turkish Journal of Agricultural Research*, 7(3), 246-252. https://doi.org/10.19159/tutad.685342

Baştürk, M.Ö., Turgut, K. and Hocaoğlu, A.K. (2021). *Görüntü işleme tabanlı elma ağacinda rekolte tahmini*. Paper presented at the Union Radio-Scientifique Internationale. Kocaeli, Turkey. Retrieved from http://ursitr2021.gtu.edu.tr/MCMSR/papers/URSI-TR_2020_paper_84.pdf

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris D. and Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11), 3758-3813. https://doi.org/10.3390/s21113758

Bregaglio, S., Fischer K., Ginaldi, F., Valeriano, T. and Giustarini, L. (2021). The HADES yield prediction system–a case study on the Turkish hazelnut sector. *Frontiers in Plant Science*, 12, 1-14. https://doi.org/10.3389/fpls.2021.665471

Charoen-Ung, P. and Mittrapiyanuruk, P. (2018). Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning. In H. Unger, S. Sodsee and P. Meesad (Eds), *Recent Advances in Information and Communication Technology 2018* (pp. 33-42). Paper Presented at International Conference on Computing and Information Technology, Cham: Springer. https://doi.org/10.1007/978-3-319-93692-5_4

Chen, P., Xie, G., Liu, H., Liang, L., Gao H., Wang, D. and Ji, W.J. (2020). *Online output estimation for multimode process with dynamic time-delay*. Paper presented at the 39th Chinese Control Conference (CCC). Shenyang, China. doi:10.23919/CCC50068.2020.9189279

Chlingaryan, A., Sukkarieh, S. and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61-69. https://doi.org/10.1016/j.compag.2018.05.012

Dang, C., Liu, Y., Yue, H., Qian, J. and Zhu, R. (2021). Autumn crop yield prediction using data-driven approaches: Support vector machines, random forest, and deep neural network methods. *Canadian Journal of Remote Sensing*, 47(2), 162-181. https://doi.org/10.1080/07038992.2020.1833186

Ercan, Ş., Öztep, R., Güler, D. and Saner, G. (2019). Tarım 4.0 ve Türkiye'de uygulanabilirliğinin değerlendirilmesi. *Tarım Ekonomisi Dergisi*, 25(2), 259-265. https://doi.org/10.24181/tarekoder.650762

Everingham, Y., Sexton, J., Skocaj, D. and Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, 36(27), 1-9. https://doi.org/10.1007/s13593-016-0364-z

FAO. (2021). *The state of food and agriculture 2021: Making agrifood systems more resilient to shocks and stresses*. Rome: FAO. https://doi.org/10.4060/cb4476en

Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D., Pozza, L.E., Ugbaje, S.U., . . . and Bishop, T.F. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, 20, 1015-1029. https://doi.org/10.1007/s11119-018-09628-4

França, T., Martins, A., Braga, B. and Ayala, H.V. (2022). Feature engineering to cope with noisy data in sparse identification. *Expert Systems with Applications*, 188, 115995. https://doi.org/10.1016/j.eswa.2021.115995

Gandhi, N., Armstrong L.J., Petkar O. and Tripath, A.K. (2016). *Rice crop yield prediction in India using support vector machines*. Paper presented at the 13th International Joint Conference on Computer Science and Software Engineering (JCSSE). Khon Kaen, Thailand. doi:10.1109/JCSSE.2016.7748856

Gopal, P.M. and Bhargavi, R. (2019a). Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. *Applied Artificial Intelligence*, 33(7), 621-642. https://doi.org/10.1080/08839514.2019.1592343

Gopal, P.M. and Bhargavi, R. (2019b). A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*, 165, 104968. https://doi.org/10.1016/j.compag.2019.104968

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., . . . and Kim, S.–H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, 11(6), 1-15. https://doi.org/10.1371/journal.pone.0156571

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C. and Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US midwest. *Environmental Research Letters*, 15, 064005. https://doi.org/10.1088/1748-9326/ab7df9

Kawasaki, K. and Lichtenberg, E. (2015). *Quality versus quantity effects of pesticides: Joint estimation of quality grade and crop yield*. Paper presented at the Agricultural and Applied Economics Association (AAEA) Conferences. San Francisco, USA. doi:10.22004/ag.econ.204848

Kaya, Y. and Polat, N. (2021). Wheat yield estimation using vegetation indices. *Dicle Üniversitesi Mühendislik Dergisi*, 12(1), 99-110. https://doi.org/10.24012/dumf.860325

Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10, 621. https://doi.org/10.3389/fpls.2019.00621

Khanal, S., Fulton, J., Klopfenstein A., Douridas, N. and Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, 153, 213-225. https://doi.org/10.1016/j.compag.2018.07.016

Khosla, E., Dharavath, R. and Priya, R. (2020). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. *Environment, Development and Sustainability*, 22, 5687-5708. https://doi.org/10.1007/s10668-019-00445-x

Kılavuz, E. and Erdem, İ. (2019). Agriculture 4.0 applications in the world and transformation of Turkish agriculture. *Social Sciences*, 14(4), 133-157. http://dx.doi.org/10.12739/NWSA.2019.14.4.3C0189

Kırmıkıl, M. and Ertaş, B. (2020). A sustainable future with agriculture 4.0. *Icontech International Journal*, 4(1), 1-12. https://doi.org/10.46291/ICONTECHvol4iss1pp1-12

Klompenburg, T.V., Kassahun, A. and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 17, 105709. https://doi.org/10.1016/j.compag.2020.105709

Lamichhane, J.R. (2017). Pesticide use and risk reduction in European farming systems with IPM: An introduction to the special issue. *Crop Protection*, 97, 1-6. doi:10.1016/j.cropro.2017.01.017

Leo, S., Migliorati, M.D. and Grace, P.R. (2020). Predicting within-field cotton yields using publicly available datasets and machine learning. *Agronomy Journal*, 113(2), 1150-1163. https://doi.org/10.1002/agj2.20543

Liakos, K.G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674-2703. https://doi.org/10.3390/s18082674

Lin, A.Y., Zhang, M. and Selpi, S. (2018). Using scaling methods to improve support vector regression's performance for travel time and traffic volume predictions. In I. Rojas, H. Pomares and O. Valenzuela (Eds.), *Time series analysis and forecasting ITISE 2017 contributions to statistics* (pp. 115-127). Paper Presented at International Work-Conference on Time Series Analysis, Cham: Springer. https://doi.org/10.1007/978-3-319-96944-2_8

Lischeid, G., Webber, H., Sommer, M., Nendel, C. and Ewert, F. (2022). Machine learning in crop yield modelling: A powerful tool, but no surrogate for science. *Agricultural and Forest Meteorology*, 312, 108698. https://doi.org/10.1016/j.agrformet.2021.108698

Liu, Y., Miller, E. and Habib, K.N. (2021). Detecting transportation modes using smartphone data and GIS information: Evaluating alternative algorithms for an integrated smartphone-based travel diary imputation. *Transportation Letters*, 1958591. https://doi.org/10.1080/19427867.2021.1958591

Lobell, D.B. and Burke, M.B. (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environmental Research Letters*, 3(3), 034007. https://doi.org/10.1088/1748-9326/3/3/034007

McQueen, R.J., Gamer, S.R., Nevill-Manning, C.G. and Witten, I.H. (1995). Applying machine learning to agricultural data. *Computers and Electronics in Agriculture*, 12, 275-293. https://doi.org/10.1016/0168-1699(95)98601-9

Millán-Castillo, R.S., Morgado, E. and Goya-Esteban, R. (2020). On the use of decision tree regression for predicting vibration frequency response of handheld probes. *IEEE Sensors Journal*, 20(8), 4120 - 4130. https://doi.org/10.1109/JSEN.2019.2962497

OECD/FAO. (2020). *OECD-FAO agricultural outlook 2020-2029*. Retrieved from https://doi.org/10.1787/1112c23b-en

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144(1), 31-43. https://doi.org/10.1017/S0021859605005708

Ozdogan, B., Gacar, A. and Aktas, H. (2017). Digital agriculture practices in the context of agriculture 4.0. *Pressacademia*, 4, 184-191. https://doi.org/10.17261/Pressacademia.2017.448

Pant, J., Pant, R., Singh, M.K., Singh, D.P. and Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, 46, 10922-10926. https://doi.org/10.1016/j.matpr.2021.01.948

Patrício, D.I. and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153, 69-81. https://doi.org/10.1016/j.compag.2018.08.001

Paudel, D., Boogaard, H. Wit, A.D., Janssen, S., Osinga, S., Pylianidis, C. and Athanasiadis, I.N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 187, 103016. https://doi.org/10.1016/j.agsy.2020.103016

Paudel, D., Boogaard, H., Wit, A.D, Velde, M.D., Claverie, M., . . . Athanasiadis, I.N. (2022). Machine learning for regional crop yield forecasting in Europe. *Field Crops Research*, 276, 108377. https://doi.org/10.1016/j.fcr.2021.108377

Pekel, E. (2020). Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology*, 139, 1111–1119. https://doi.org/10.1007/s00704-019-03048-8

PSB. (2019). *Eleventh development plan (2019-2023).* Retrieved from https://www.sbb.gov.tr/

Rahman, M.M., Haq N. and Rahman, R.M. (2014). *Machine learning facilitated rice prediction in Bangladesh*. Paper presented at the 2014 Annual Global Online Conference on Information and Computer Technology. Louisville, USA. Retrieved from https://ieeexplore.ieee.org/

Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A. and Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE Access*, 9, 63406-63439. https://doi.org/10.1109/ACCESS.2021.3075159

Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P. and Ciampitti, I.A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology,* 284, 107886. https://doi.org/10.1016/j.agrformet.2019.107886

Scornet, E., Biau, G. and Vert, J.-P. (2015). Consistency of random forests. *The Annals of Statistics*, 43(4), 1716-1741. https://doi.org/10.1214/15-AOS1321

Shah, A., Dubey, A., Hemnani, V., Gala, D. and Kalbande, D.R. (2018). Smart farming system: Crop yield prediction using regression techniques. In H. Vasudevan, A. Deshmukh and K. Ray (Eds.), *Proceedings of International Conference on Wireless Communication* (pp. 49-56). Paper Presented at the International Conference on Wireless Communication, Singapore: Springer. https://doi.org/10.1007/978-981-10-8339-6_6

Shi, X., An, X., Zhao, Q., Liu, H., Xia, L., Sun, X. and Guo, Y. (2019). State-of-the-art internet of things in protected agriculture. *Sensors*, 19(8), 1833. https://doi.org/10.3390/s19081833

Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S. and Singh, A.K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. *PLoS ONE*, 16(6), e0252402. https://doi.org/10.1371/journal.pone.0252402

Şimşek, O., Mermer, A., Yıldız, H., Özaydın, K.A. and Çakmak, B. (2007). Estimation of wheat yield for Turkey using AgroMetShell model. *Journal of Agricultural Sciences*, 13(3), 299-307. Retrieved from https://dergipark.org.tr/en/pub/ankutbd/

Tang, S., Zhu, Q., Zhou, X., Liu, S. and Wu, M. (2002). *A conception of digital agriculture*. Paper presented at the International Geoscience and Remote Sensing Symposium. Toronto, Canada. https://doi.org/10.1109/IGARSS.2002.1026858

Tauger, M.B. (2011). *Agriculture in world history*. Oxfordshire: Routledge.

Toscano, N.C., Sances, F.V., Johnson, M.W. and Lapre, L.F. (1982). Effect of various pesticides on lettuce physiology and yield. *Journal of Economic Entomology*, 75(4), 738-741. https://doi.org/10.1093/jee/75.4.738

Trnka, M., Olesen, J.E., Kersebaum, K.C., Rötter, R.P., Brázdil, R., Eitzinger, J., . . . Semerádová, D. (2016). Changing regional weather-crop yield relationships across Europe between 1901 and 2012. *Climate Research*, 70, 195-214. https://doi.org/10.3354/cr01426

Vaid, K. and Ghose, U. (2020). Predictive analysis of manpower requirements in scrum projects using regression technique. *Procedia Computer Science*, 173, 335–344. https://doi.org/10.1016/j.procs.2020.06.039

Vanli, Ö., Ahmad, I. and Ustundag, B.B. (2020). Area estimation and yield forecasting of wheat in southeastern Turkey using a machine learning approach. *Journal of the Indian Society of Remote Sensing*, 48(21), 1757-1766. https://doi.org/10.1007/s12524-020-01196-3

Varjovi, M.H. and Talu, M.F. (2016). *Kayısı için otomatik rekolte tahmin sistemi*. Paper presented at International Conference on Artificial Intelligence and Data Processing. Malatya, Turkey. Retrieved from https:// http://idap.inonu.edu.tr/

Washuck, N., Hanson, M. and Prosser, R. (2022). Yield to the data: Some perspective on crop productivity and pesticides. *Pest Management Science*, 78(5), 1765-1771. https://doi.org/10.1002/ps.6782

Xie, S., Feng, H., Yang, F., Zhao, Z., Hu, X., Wei, C., . . . Geng, Y. (2019). Does dual reduction in chemical fertilizer and pesticides improve nutrient loss and tea yield and quality? A pilot study in a green tea garden in Shaoxing, Zhejiang Province, China. *Environmental Science and Pollution Research*, 26, 2464–2476. https://doi.org/10.1007/s11356-018-3732-1

Xu, M., Watanachaturaporn, P., Varshney, P.K. and Arora, M.K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97, 322-336. https://doi.org/10.1016/j.rse.2005.05.008

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., . . . Wu, X. (2019). Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. *Ecological Indicators*, 101, 943-953. https://doi.org/10.1016/j.ecolind.2019.01.059

Zambon, I., Cecchini, M., Egidi, G., Saporito, M.G. and Colantoni, A. (2019). Revolution 4.0: Industry vs. agriculture in a future development for SMEs. *Processes*, 7(36), 1-16. https://doi.org/10.3390/pr7010036

Zarei, A.R., Mahmoudi, M.R., Shabani, A. and Achite, M. (2021). Determination of the most important meteorological parameters affecting the yield and biomass of barley and winter wheat using the random forest algorithm. *Paddy and Water Environment*, 19, 199-216. https://doi.org/10.1007/s10333-020-00832-5

Zuo, X., Guo, H., Shi, S. and Zhang, X. (2020). Comparison of six machine learning methods for estimating PM2.5 concentration using the himawari-8 aerosol optical depth. *Journal of the Indian Society of Remote Sensing*, 48(9), 1277–1287. https://doi.org/10.1007/s12524-020-01154-z