Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi
Journal of Applied Sciences of Mehmet Akif Ersoy University

Determination of Important Variables in Food Security Classification Using Random Forest[1]

Özlem EŞTÜRK[1*]

[1] Asst. Prof. Dr., Ardahan University, Faculty of Economics and Administrative Sciences, Department of Economics, Ardahan, Türkiye

Geliş Tarihi/Received: 19.12.2021
Kabul Tarihi/Accepted: 23.02.2022
Doi:10.31200/makuubd.1038467
Araştırma Makalesi/Research Article

ABSTRACT

Seasonal agricultural workers are the most disadvantaged group of work forces in terms of poverty even though they are significant contributors to the agricultural economy in Turkey. The objectives of this study were to determine the food security status of seasonal agricultural workers and to determine the most important variables in the classification of household food security status for the seasonal agriculture workers. Responses of seasonal apricot workers in Malatya to 18 questions of the Household Food Security Survey Module (HFSSM) were analyzed using the Random Forests (RF) algorithm (n = 65). Results indicated that 55.4% of households suffered from food insecurity, where 7.7% of them with moderate hunger and 13.8% of them with severe hunger. The area under curve value of the RF model was estimated at 0.846 as the classification accuracy. The question "running out of food before having money to buy more" was the most important variable in the classification of the food security groups. Seasonal agricultural workers are prone to food insecurity and poverty due to low income and job insecurity in Turkey and in the world. Therefore, it is important to implement social aid programs to solve food insecurity issue in risk groups like seasonal agricultural workers.

Keywords: Food Security, Random Forest, HFSSM, Seasonal Agricultural Worker.

[1] This study was approved by Ardahan University Ethics Committee on 07.01.2022 with the decision number of 67796128-000-2200000871.

* Sorumlu yazar/Corresponding author
E-mail/e-ileti: ozlemesturk@ardahan.edu.tr

Gıda Güvencesi Düzeyi Sınıflandırılmasında Kullanılan Önemli Göstergelerin Random Forest Yöntemine Göre Belirlenmesi

ÖZET

Mevsimlik tarım işçileri, Türkiye'de tarım ekonomisine önemli katkılar sağlasalar da, yoksulluk açısından en dezavantajlı işgücü grubudur. Bu çalışmada mevsimlik tarım işçilerinin hane halkı gıda güvencesi durumlarının belirlenmesi ve gıda güvencesi sınıflandırmasına etki eden en önemli değişkenlerin belirlenmesi hedeflenmiştir. Malatya'da mevsimlik kayısı işçilerinin 18 sorudan oluşan Hanehalkı Gıda Güvencesi Anket Modülü (HFSSM) sorularına verdikleri yanıtlar Random Forest (RF) algoritması ile analiz edilmiştir (n = 65). Sonuçlar, hanehalklarının %55.4'ünün gıda güvencesiz olduğunu, bunlardan %7.7'sinin orta düzeyde, %13.8'inin ise ciddi düzeyde açlığın olduğu gıda güvencesizliği durumu yaşadıklarını göstermiştir. RF modelinin sınıflandırma doğruluğunu gösteren eğri altındaki alan değeri, 0.846 olarak tahmin edilmiştir. Gıda güvencesi gruplarının sınıflandırılmasında en önemli değişken "Daha fazla gıda alacak paraya sahip olmadan gıdanın bitmesi" sorusu olmuştur. Mevsimlik tarım işçileri, Türkiye'de ve dünyada düşük gelir ve iş güvencesizliği nedeniyle gıda güvencesizliği ve yoksulluk tehdidi altındadır. Bu nedenle mevsimlik tarım işçileri gibi risk gruplarında gıda güvencesizliği probleminin çözümüne yönelik sosyal yardım programlarının uygulanması önemlidir.

Anahtar kelimeler: Gıda Güvencesi, Random Forest, HFSSM, Mevsimlik Tarım İşçisi.

1. INTRODUCTION

Seasonal agricultural workers are defined as temporary or travelling employees during any stage of production in farmlands in return for a wage with or without a contract (Fereli vd., 2016). They are employed in spring and summer during which agricultural production intensifies. Seasonal agricultural workers are mostly concentrated in Cukurova, Aegean, Marmara and Black Sea regions of Turkey (Öz & Bulut, 2013). Seasonal agricultural workers, play a significant role in agricultural production of Turkey, face serious problems such as inadequate working and living conditions including low wages and unsanitary conditions, and lack of social security and food security.

Food security can be described as having a sustainable access to adequate and healthy food (FAO, 2002). As it was recognized by the universal declaration of human rights of United Nations in 1948, all people and nations have the right for food security. According to the Global

Food Security Index, Turkey ranked in the 48th with a score of 64.1 among 113 countries (Anonymous, 2021). Although the quantity of food does not usually constitute a problem, malnutrition (deficiencies, or excesses in a person's intake of nutrients) is widespread, in particular, among the seasonal agricultural workers in Turkey.

Apricot is an important crop for Turkey. Globally, Turkey is the leading apricot producer, accounting for 23.14% (985 thousand tons) of the total production (FAOSTAT, 2017). In particular, Malatya produces 67.27% of the fresh apricots grown in Turkey (TURKSTAT, 2019).

There are numerous studies in literature establishing the relationship between household food insecurity and a variety of physical (Gucciardi vd., 2009; Vozoris & Tarasuk, 2003) and mental (Davison & Kaplan, 2015; Heflin vd., 2005; Martin vd., 2016) health problems. Payne-Sturges vd. (2017) investigated the food insecurity in college students and find out 15% of the students were food insecure and 16% were at risk of food insecurity. According to this survey, food secure students were less likely to report depression symptoms compared to the food insecure or food security at-risk students.

In general, questionnaire-based datasets suffer from a small sample size, outliers, and the lack of normal distribution. Thus, non-parametric data analyses such as decision trees play a significant role in deriving information and knowledge from such data. In terms of classification and prediction accuracy, recent advances have been made in the development of efficient decision tree algorithms such as bagging, boosting, and bootstrapping (L. Breiman, 1996; Freund & Schapire, 1996). The random forests (RF) algorithm introduced by Leo Breiman (2001) has been proved to be one of the most powerful machine learning methods. Even its stochastic characteristic, resistance to outliers, and/or missing values and interpretability in terms of variable importance have offset its black-box nature. However, there still exists a large gap about the application of RF to socio-economic survey data.

To the best of our knowledge, there is no study about the machine learning-based quantification and assessment of household food security in related literature. Therefore, the objectives of the present study were to (1) quantify the status of household food security of seasonal apricot workers in Malatya by using the Household Food Security Survey Module (HFSSM) developed by the U.S. Department of Agriculture (USDA), and (2) assess the most important variables in the determination of household food security by using the RF algorithm.

2. MATERIAL AND METHODS

2.1. Data Collection

The study was conducted between July and August 2018 for seasonal apricot workers in Malatya, Turkey. Randomly selected participants (n = 65) were over age of 18 migrated from other provinces to work as a seasonal agricultural worker. Households were also evaluated for 37 demographic variables such as gender, age, occupation, job status, income and food consumption habits. The survey consisted of 18 questions intended to measure the prevalence and severity of food (in)security in households (Nord vd., 2008). All the questions were referenced to the previous 12 months. The food security status of each household was assessed using their responses to 10 questions for households without child and to 18 questions for households with child. In the survey, the responses of "often" or "sometimes" were coded as affirmative (1), and "never" was coded as negative (0), while "yes", and "no" were coded as 1 and 0, respectively. A score about household food security was calculated from the number of affirmative responses to the questions for the two types (without child versus with child). Based on the overall score, households were classified into the food security categories (Table 1).

Table 1. Scoring of household food security scales

Food security level	Number of affirmative responses	Group
Food secure	0	A
Food security at-risk	1-2	B
Food insecurity without hunger	3-5	C
Food insecurity moderate hunger	6-8	D
Food insecurity severe hunger	≥ 9	E

The questionnaire form used in the research was approved by the Ardahan University Ethics Committee.on 07.01.2022 by the decision number of E-67796128-000-2200000871.

2.2. Random Forests Classification

In this study, variable importance scores in the classification of household food security groups of the seasonal agricultural workers were determined using the RF method. The RF classification algorithm is based on the growth and voting of the ensemble of multiple decision trees (forest) without trimming via the bootstrapping technique. Bootstrapping is a randomly resampling technique with replacement. The model performance is measured using out-of-bag

(OOB) error rate, an internal validation process. Each tree is weighted based on the OOB error rate, with the decision tree (model) with the lowest error rate having the highest weight and subjected to a voting process for its class prediction. Overall, the RF algorithm as an ensemble method reflects the outcome of these weighted votes. The statistical modeling was performed using Salford Predictive Modeler (R) software 8.3.0 (Salford Systems, San Diego, California, USA).

3. RESULTS AND DISCUSSION

According to the survey data, 55.4% of the seasonal apricot workers suffered from food insecurity, where 7.7% of them were with moderate hunger and 13.8% of them were with severe hunger. Wirth vd. (2007) investigated food security status of seasonal agricultural workers (n=102) in North Carolina in 2004 and reported that while 52.9 of households were food secure, 32.4% of them were food insecure.

The number of decision trees grown in RF was chosen as 1 000, as suggested by Breiman (2004). No separate test data other than the internal validation data of OOB were used in the RF classification model. The RF model error rate is presented in Figure 1. The overall error rate of the RF model was estimated at 0.207 when n = 1 000 decision trees (Table 2). The area under curve (AUC) values are the measure of the goodness of fit of the RF models (Leo Breiman, 2001; Evans vd., 2011; Grossmann vd., 2010). The AUC values higher than 0.90, between 0.90 and 0.81, 0.80 and 0.71, 0.70 and 0.61, and lower than 0.60 were refer to excellent, good, acceptable, weak and failing models, respectively, in terms of the model fit (Özdemir, 2018; Süel, 2014). In our study, the AUC value of the RF model was estimated at 0.846 (Table 2).

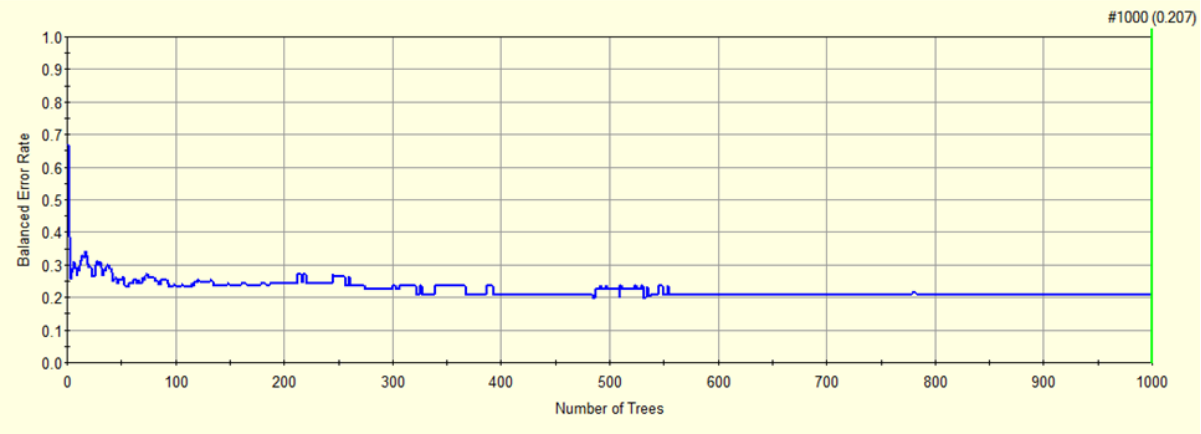

Figure 1. Change in error rate with the increased number of decision trees in RF model

The correct and false classifications of the RF model were given in the Table 3. The classification accuracy was perfect (100%) for the group B (food security at-risk). It was relatively low for the group C (food insecure without hunger) and group D (food insecure with moderate hunger) and good for the group E (food insecure with severe hunger), where correct classifications were 15 out of 22, 3 out of 5 and 8 out of 9, respectively. There was no food secure household (group A) among the seasonal agricultural workers survey conducted. The overall correct classification rate of the RF model was 84.62%.

Table 2. Model error measures

Name	**OOB**
Balanced error rate (Simple average over classes)	0.20732
Classification accuracy (Baseline threshold)	0.84615

Table 3. Confusion matrix-OOB

Actual Class	Total Class	Percent Correct	Predicted Classes			
			2 N = 32	3 N =16	4 N = 8	5 N = 9
2	29	100.00%	29	0	0	0
3	22	68.18%	3	15	4	0
4	5	60.00%	0	1	3	1
5	9	88.89%	0	0	1	8
Total:	65					
Average:		79.27%				
Overall % Correct:		84.62%				

The variable importance was estimated using the permutation and Gini methods (Table 4). According to the Gini method, the most important variables in the classification of the household food security groups in the following order of decreasing importance were Q2 (running out of food before having money to buy more), Q5-1 (frequency of household adults cutting the size of meals or skipping meals because of insufficient money for food), Q4-1 (frequency of eating less food because of insufficient money), Q5 (cutting the size of meals or

skipping meals because of insufficient money), Q1 (worrying about running out of food before having money to buy more), Q4 (eating less food because of insufficient money), S1 (number of person in a household), S6D (frequency of meat and meat products consumption), CH3 (children not eating enough because of not being able to afford enough food) and Q8 (losing weight because of insufficient money).

Table 4. Gini variable importance scores

Variable	*Score*	
Q2	100.00	
Q51	88.15	
Q41	75.19	
Q5	73.06	
Q1	66.10	
Q4	65.58	
S1	53.34	
S6D_	48.19	
CH3	44.10	
Q8	43.06	
CH6	36.20	
S3	36.19	
S6F_	35.27	
S5	34.20	
Q6	33.39	
CH4	29.29	
CH1	28.66	
S6E_	26.48	
S4	24.06	
S2D	21.37	
CH5	20.47	
S2A_	18.99	
S6A_	16.95	
Q3	14.65	
S6I_	12.33	
S6H	6.90	
S2B_	6.80	
S6C	5.81	
CH2	5.21	
S2F_	4.41	
S6B_	3.75	
CH51	3.07	
Q7	2.28	
Q71	1.50	

The comparison of the permutation (Table 5) and Gini methods showed that the first two most important variables of Q2 and Q5-1 were in the same order. The others based on the permutation method were Q1, Q5, Q8, Q4-1, Q4, S6D, CH3, and S3 (marital status), respectively.

Godrich vd. (2019) investigated the differences in self-esteem and self-efficacy for healthy lifestyle choices between children living in food secure and food insecure households and reported that household income was a significant determinant underlying the association of food insecurity with self-esteem and self-efficacy.

Table 5. Permutation variable importance scores

<table>
<tr><th>Variable</th><th>Score</th><th></th></tr>
<tr><td>Q2</td><td>100.00</td><td>||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>Q51</td><td>42.60</td><td>||||||||||||||||||||</td></tr>
<tr><td>Q1</td><td>40.80</td><td>|||||||||||||||||||</td></tr>
<tr><td>Q5</td><td>40.70</td><td>|||||||||||||||||||</td></tr>
<tr><td>Q8</td><td>39.44</td><td>||||||||||||||||||</td></tr>
<tr><td>Q41</td><td>37.49</td><td>|||||||||||||||||</td></tr>
<tr><td>Q4</td><td>34.52</td><td>||||||||||||||||</td></tr>
<tr><td>S6D_</td><td>31.40</td><td>|||||||||||||||</td></tr>
<tr><td>CH3</td><td>31.22</td><td>||||||||||||||</td></tr>
<tr><td>S3</td><td>24.22</td><td>|||||||||||</td></tr>
<tr><td>Q6</td><td>23.28</td><td>|||||||||||</td></tr>
<tr><td>CH1</td><td>21.91</td><td>||||||||||</td></tr>
<tr><td>S6F_</td><td>19.78</td><td>|||||||||</td></tr>
<tr><td>CH6</td><td>17.19</td><td>||||||||</td></tr>
<tr><td>CH4</td><td>14.10</td><td>||||||</td></tr>
<tr><td>CH5</td><td>9.23</td><td>||||</td></tr>
<tr><td>S2A_</td><td>6.83</td><td>|||</td></tr>
<tr><td>S1</td><td>6.43</td><td>|||</td></tr>
<tr><td>S6A_</td><td>2.67</td><td>|</td></tr>
<tr><td>S5</td><td>2.54</td><td>|</td></tr>
<tr><td>S6E_</td><td>1.97</td><td></td></tr>
<tr><td>S6I_</td><td>1.97</td><td></td></tr>
<tr><td>CH2</td><td>1.64</td><td></td></tr>
<tr><td>Q3</td><td>1.37</td><td></td></tr>
<tr><td>S4</td><td>1.09</td><td></td></tr>
<tr><td>S2C</td><td>1.09</td><td></td></tr>
<tr><td>S2B_</td><td>1.06</td><td></td></tr>
<tr><td>CH51</td><td>0.34</td><td></td></tr>
<tr><td>S2D</td><td>0.20</td><td></td></tr>
</table>

4. CONCLUSIONS

Results indicated that seasonal apricot workers suffered from food insecurity. Food security was largely dependent on the purchasing power of food. The AUC value of the RF classification model (0.846) indicated its utility in the detection of the driving forces and causation patterns behind household food (in)security of the seasonal agricultural workers in Turkey and in the world. The findings of this study showed how the families of agricultural workers perceiving food security and necessity of developing different intervention strategies for different populations. Thus, public services addressing the chronic food insecurity status of

families of seasonal agricultural workers and reducing the stress level as a result of food insecurity are important. Governments should put food security programs into place for the households like seasonal agricultural workers to solve food insecurity problem in these groups.

Abbreviations and Symbols	
Q1-Q8	HFSSM survey questions
CH1-CH6	HFSSM survey questions for households with child
S1-S6	Demographic survey question

CONFLICT STATEMENT

Author have declared no conflict of interest.

REFERENCES / KAYNAKLAR

Anonymous. (2021). *Global food security index New York, USA*. https://foodsecurityindex.eiu.com/

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140. doi:10.1023/a:1018054314350

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/a:1010933404324

Breiman, L. (2004). Consistency for a simple model of random forests (Technical Report 670). California: University of California at Berkeley, https://www.stat.berkeley.edu/~breiman/RandomForests/consistencyRFA.pdf

Davison, K., & Kaplan, B. (2015). Food insecurity in adults with mood disorders: Prevalence estimates and associations with nutritional and psychological health. *Annals of General Psychiatry, 14*. doi:10.1186/s12991-015-0059-x

Evans, J. S., Murphy, M. A., Holden, Z. A., & Cushman, S. A. (2011). Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology* (pp. 139-159). New York: Springer.

FAO. (2002). *The state of food insecurity in the world 2001*. Erişim tarihi: 14.03.2021, http://www.fao.org/3/y1500e/y1500e00.htm

FAOSTAT. (2017). *FAO statistics*. Erişim tarihi: 15.03.2021, http://www.fao.org/faostat

Fereli, S., Aktaç, Ş., & Güneş, F. E. (2016). Working conditions, nutritional status and problems seen on seasonal agricultural workers. *Gazi Üniverstesi Sağlık Bilimleri Dergisi, 1*(3), 36-47.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Paper presented at the Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, Bari, Italy.

Godrich, S., K. Loewen, O., Blanchet, R., Willows, N., & Veugelers, P. (2019). Canadian children from food insecure households experience low self-esteem and self-efficacy for healthy lifestyle choices. *Nutrients, 11*, 675. doi:10.3390/nu11030675

Grossmann, E., Ohmann, J., Kagan, J., May, H., & Gregory, M. (2010). Mapping ecological systems with a random forest model: Tradeoffs between errors and bias. *Gap Analysis Bulletin, 17*(1), 16-22.

Gucciardi, E., Vogt, J. A., DeMelo, M., & Stewart, D. E. (2009). Exploration of the relationship between household food insecurity and diabetes in Canada. *Diabetes care, 32*(12), 2218-2224. doi:10.2337/dc09-0823

Heflin, C. M., Siefert, K., & Williams, D. R. (2005). Food insufficiency and women's mental health: Findings from a 3-year panel of welfare recipients. *Social Science & Medicine, 61*(9), 1971-1982. doi:https://doi.org/10.1016/j.socscimed.2005.04.014

Martin, M. S., Maddocks, E., Chen, Y., Gilman, S. E., & Colman, I. (2016). Food insecurity and mental illness: Disproportionate impacts in the context of perceived stress and social isolation. *Public Health, 132*, 86-91. doi:https://doi.org/10.1016/j.puhe.2015.11.014

Nord, M., Andrews, M., & Carlson, S. (2008). Household food security in the United States, 2008 (Economic Research Report No. 83). Washington, DC, US. https://www.hsdl.org/?view&did=31871.

Öz, C. S., & Bulut, E. (2013). The status of seasonal agricultural workers in Turkish legislation. *Labour World, 1*(1), 94-111.

Özdemir, S. (2018). Potential Distribution Modelling and mapping using Random Forest method: An example of Yukarıgökdere distric. *Turkish Journal of Forestry, 19*(1), 51-56.

Payne-Sturges, D., Tjaden, A., Caldeira, K., Vincent, K., & Arria, A. (2017). Student hunger on campus: Food insecurity among college students and implications for academic institutions. *American Journal of Health Promotion, 32*(2), 349-354. doi:10.1177/0890117117719620

Süel, H. (2014). *Mapping habitat suitability of game animals in Sütçüler district, Isparta* (Phd dissertation). Suleyman Demirel University, Isparta.

TURKSTAT. (2019). *Crop production statistics*. Erişim tarihi: 02.04.2020, http://www.turkstat.gov.tr/

Vozoris, N., & Tarasuk, V. (2003). Household food insufficiency is associated with poorer health. *The Journal of Nutrition, 133*, 120-126. doi:10.1093/jn/133.1.120

Wirth, C., Strochlic, R., & Getz, C. (2007). *Hunger in the fields: Food insecurity among farmworkers in Fresno county*. California Institute for Rural Studies.