Acta Infologica
acin 2023, 7 (2), 243–252
DOI: 10.26650/acin.1251650

Research Article / Araştırma Makalesi

Determining the Happiness Class of Countries with Tree-Based Algorithms in Machine Learning

Makine Öğrenmesinde Ağaç Tabanlı Algoritmalarla Ülkelerin Mutluluk Sınıfının Belirlenmesi

Merve Doğruel[1], Selin Soner Kara[2]

[1](Assist. Prof.), University of Istanbul Esenyurt, Faculty of Business and Management Sciences, Department of Management Information Systems, Istanbul, Turkiye
[2](Prof. Dr.), Yildiz Technical University, Faculty of Mechanical Engineering, Department of Industrial Engineering, Istanbul, Turkiye

Corresponding author : Merve DOĞRUEL
E-mail : mervedogruel@esenyurt.edu.tr

ABSTRACT

Today, the concept of happiness is a frequently researched subject in the fields of economy, medicine, and social and political fields, as well as psychology. It has been an important research area for everyone, from policymakers to companies, to determine the factors affecting happiness. With machine learning algorithms, it is possible to make classifications with very high accuracy. The aim of this study is to use tree-based machine learning algorithms to classify the happiness scores of countries. In order to accomplish this, data from the World Happiness Index published in 2022 were used. On these data, tree-based algorithms CART, tree-based ensemble algorithms Bagging, and Random Forest were used. The test data of the model were obtained with 85% precision, recall, and F1 metrics, which were calculated using Bagging and Random Forest algorithms. The outcomes of the models obtained during the study were interpreted.

Keywords: Machine learning, World Happiness Index, Ensemble learning

ÖZ

Mutluluk kavramı günümüzde psikoloji alanı dışında ekonomi, tıp, sosyal ve politik alanlarda da sıklıkla araştırılan bir konu haline gelmiştir. Mutluluğu etkileyen faktörlerin belirlenmesi, politika yapıcılardan işletmelere kadar önemli bir araştırma alanı olmuştur. Makine öğrenmesi algoritmaları ile yüksek doğrulukta sınıflandırmalar çalışmaları yapmak mümkündür. Bu çalışmada, ağaç tabanlı makine öğrenmesi algoritmaları kullanılarak ülkelerin mutluluk puanlarının sınıflandırılması amaçlanmaktadır. Bu amaçla 2022 yılında yayınlanan Dünya Mutluluk Endeksi'nden alınan veriler kullanılmıştır. Bu veriler üzerinde ağaç tabanlı algoritmalar SRT, ağaç tabanlı topluluk algoritmaları torbalama ve rastgele orman kullanılmıştır. Torbalama ve rastgele orman algoritmaları ile elde edilen modelin test verilerinde %85 kesinlik, duyarlılık ve F1 metrikleri hesaplanmıştır. Çalışmada elde edilen bu modellerin sonuçları yorumlanmıştır.

Anahtar Kelimeler: Makine öğrenmesi, Dünya Mutluluk Endeksi, Topluluk öğrenmesi

Submitted	: 15.02.2023
Revision Requested	: 26.07.2023
Last Revision Received	: 01.08.2023
Accepted	: 01.08.2023
Published Online	: 24.08.2023

1. INTRODUCTION

Happiness was previously only discussed in the field of psychology, since it is a subjective concept that cannot be defined with an absolute definition. Over time, the concept of happiness, which is affected by many factors, has become an important research topic for economics under the heading economics of happiness (Öztürk & Suluk, 2020), which is separate from the fields of medicine, sociology, and politics.

International organizations have carried out various index studies on happiness, which is important in many fields. Examples of these are the World Happiness Report, Happy Planet Index, OECD Better Life Index, and Gallup Global Emotions.

The World Happiness Report (Helliwell et al., 2022), which was first published in 2013, has been published every year and its 10th edition was published in 2022. It has a readership of over 9 million and has been cited numerous times.

When the literature studies are examined, it is seen that there are many different results regarding the factors affecting happiness and the level of importance of these factors. It is vital to determine the effects of factors affecting happiness, including the mental states of individuals, the behavior of companies, and the policy determinations of countries. For this purpose, many statistical studies have been carried out in the literature in order to determine happiness.

Farooq and Shanmugam (2022) analyzed performance metrics using validated COVID-19 datasets of countries and happiness reports showing how free citizens are. Various machine learning techniques were used, such as linear regression, logistic regression, support vector machine (SVM), Naive Bayes (NB), and K-nearest neighbor (KNN) algorithms. In their study, Kıroğlu and Yıldırım (2022) examined the determinants of happiness in Turkey using multivariate logit models. In their research, Jannahi, Sael, and Benabbou (2021) aimed to predict quality of life using basic machine learning models using data from the 2015-2021 World Happiness Index reports.

In the research conducted by Ulkhaq and Adyatama (2021), countries were clustered based on distinct clustering algorithms using the World Happiness Report 2019 data. Ibnat, Gyalmo, Alom, Abdul Awal, and Azim (2021) utilized the 2019 World Happiness Report data to identify the happiest countries and regions through supervised machine learning techniques and to assess the life satisfaction of the country. Chaudhary, Dixit, and Sahni (2020) used the World Happiness Index data from 2016-18 to build models with Predictive Modeling and Bayesian Networks methods. They then evaluated them for 2019 data.

In the study conducted by Garces, Adriatico, and Timbal (2019), the data obtained from the World Happiness Index 2014 was utilized in the cluster analysis based on the quality of life of the countries, in conjunction with the various indicators determined. Carlsen's study (2018) used the partial ranking methodology to calculate the happiness index. The author proposed a distinct ranking from the country rankings as reported in the 2016 report, arguing that the ranking obtained through this calculation method is more nuanced. Dao's study (2017) examined the direct impact of government spending on happiness in 183 countries between 1990 and 2016 using pooled OLS, fixed effects, random effects models, and cross-sectional analysis.

This study aims to establish a model with the data from the 2022 World Happiness Index report using tree-based machine learning techniques and to determine the most important indicators in the appropriate model. Machine learning applications acquire knowledge through experience, similar to humans, without direct programming. The algorithm, which learns from training data and experience, can then detect the data it encounters and perform estimation and classification with a high accuracy rate. Within the scope of the purpose of the study, the main reason for using decision trees, one of the supervised machine learning methods, is both the ease of explanation and interpretation of the results obtained and the ability of decision trees to make variable selections. Additionally, ensemble learning algorithms, which enable the creation of more than one tree, in other words, forests, by adding the concept of randomness to the decision tree algorithms, were applied within the scope of the study. Decision trees were created using the CART algorithm, and decision forests were created using the Bagging and Random Forest algorithms. Decision forests are also known as ensemble learning algorithms.

When the test data metrics showing the classification success of the decision trees obtained were examined, it was determined that the precision, recall, and F1 values were equal in the ensemble learning algorithms and higher than the CART algorithm. As in many studies in the literature, success metrics obtained with ensemble learning were higher in this study. Based on results obtained with Bagging and Random Forest ensemble learning algorithms, according to the World Happiness Index 2022 report, GDP, social support, and health life expectancy are the most important factors in determining the happiness classes of countries. In this study, world happiness classes are investigated using tree-based machine learning algorithms. It is important that policymakers who want to increase the happiness level of countries evaluate this study first.

2. TREE-BASED ALGORITHMS IN MACHINE LEARNING

The most commonly preferred decision tree in data science is a predictive model that can be expressed as the recursive division of the covariate space into subspaces. Decision trees, which were previously subject to decision theory and statistics, have been developed with applications in other fields such as time and data mining, machine learning, and pattern recognition. A sub-branch of artificial intelligence, machine learning develops a model that allows making predictions for new data by learning from training data thanks to computer software (Okumuş, Ekmekçioğlu & Kara, 2021). Decision trees are algorithms based on supervised learning in machine learning. As in other learning algorithms, the decision tree learning algorithm chosen aims to generate the most appropriate model from the training data. Afterward, the validity of the model created with the test data is tested, and if the model is approved, it is used to make predictions (Doğruel & Fırat, 2021). Decision trees are algorithms that are easy to understand and have a high success rate because they imitate human thinking ability while making decisions (Efeoğlu, 2022).

Ensemble learning is the realization of a specific learning task by creating and combining multiple models in order to arrive at a better decision than the decisions made separately. Many studies in the literature have shown that ensemble learning increases the predictive power of a single model. Ensemble learning produces more effective results, especially in decision tree models. In some sources, it is seen that the trees obtained by combining ensemble learning and decision tree learning are also called decision forests. Random Forest algorithm is the most popular decision forest (Rokach, 2016).

In this study, the CART algorithm from decision trees, Bagging and Random Forest algorithms from decision trees, and ensemble learning algorithms are employed in conjunction (decision forests).

2.1. CART

The CART (Classification and Regression Tree) algorithm is a non-parametric and non-linear decision tree algorithm that makes predictions based on repeated binary separation, used to create both classification and regression trees. If the target variable is categorical, the tree is referred to as a classification tree (CT), whereas if the target variable is continuous, the tree is referred to as a regression tree (RA) (Doğruel & Fırat, 2021).

Using the CART algorithm, a classification tree can be grown by dividing the dataset into two sub-partitions (lower leaves) on a rule-based basis by binary recursion. Each split involves a single variable; some variables can be used multiple times, while others are not utilized at all. Each sub-leaf is then further divided according to independent rules. The rule-based approach used in CART relies on a binary iterative partitioning path that divides a subset of the dataset, known as leaves, into two subsets known as sub-leaves based on minimizing a computed heterogeneity criterion. The Gini index and cross-entropy are the two most preferred heterogeneity criteria (Bel, Allard, Laurent, Cheddadi, & Bar-Hen, 2009 and James, Witten, Hastie, Tibshirani, 2013).

The Gini index is a measure of the total variance among K classes:

$$G = \sum_{k=1}^{k} \hat{P}mk(1 - \hat{P}mk) \tag{1}$$

In the formula, $\hat{P}mk$ is the proportion of training observations in the mth region that are from the kth class.

The Gini index takes a small value if all Pmk's are close to zero or one. Therefore, the Gini index is expressed as a measure of node purity, and a small Gini index indicates that a node generally contains observations from a single class. Cross-entropy is an alternative to the Gini index:

$$D = -\sum_{k=1}^{K} \hat{P}mk^{log}\hat{P}mk \tag{2}$$

Since $0 \leq \hat{P}mk \leq 1$, it follows that $0 \leq \hat{P}mk^{log}\ \hat{P}mk$. If the $\hat{P}mk$'s are all close to or close to zero, the cross-entropy will take a value close to zero. Therefore, the cross-entropy will take a small value like the Gini index, if the mth the node is pure.

Rules that include antecedents describing all nodes of the tree, from the root to the leaves, can be too complex. It may turn out that the initial decisions, especially in large trees, are not crucial for the classification of data vectors terminating in a leaf. Therefore, unnecessary rule predecessors should be removed (Grabczewski & Duch, 1999). Commonly employed pruning techniques, such as degree-based pruning of the separability of a split value (SSV) and

CART's cross-complexity pruning, rely on cross-validation. Some pre-cleaning algorithms, including cross-validation-based strategies and reduced error pruning (REP), have also been proposed as an alternative to post-pruning methods, but they are not as common as post-pruning because the results are not as good (Grabczewski, 2011).

2.2. Bagging

Working with high variance in standard decision trees can be a challenge. In this case, it is possible to obtain different results when working with different training data. In contrast, similar results will be obtained if a low variance procedure is repeatedly applied to different datasets. The process of bootstrap picking, commonly referred to as Bagging, is a general-purpose technique employed to reduce the variance of Bagging in a statistical learning method, as stated by James, Witten, Hastie, and Tibshirani (2013).

The Bagging algorithm is an ensemble learning method for creating a classifier ensemble. It involves combining basic learning algorithms trained on different samples of the training set. The basic point of the Bagging algorithm is based on the principle of providing diversity by training each basic learning algorithm that makes up the community on different training sets. Here, a simple random substitution sampling method is generally applied to generate different training sets from the data set. The results of the training sets obtained by the sampling method and the classification methods trained are merged through a majority vote (Onan, 2018)

The working principle of the Bagging algorithm is shown in Figure 1. In this principle, which is characterized by random sampling, there is no relationship between weak learning models. Randomly selected samples are subsequently returned to the data set subsequent to each extraction. This means that the previous sample can be collected continuously in the next sample (Li, 2022).

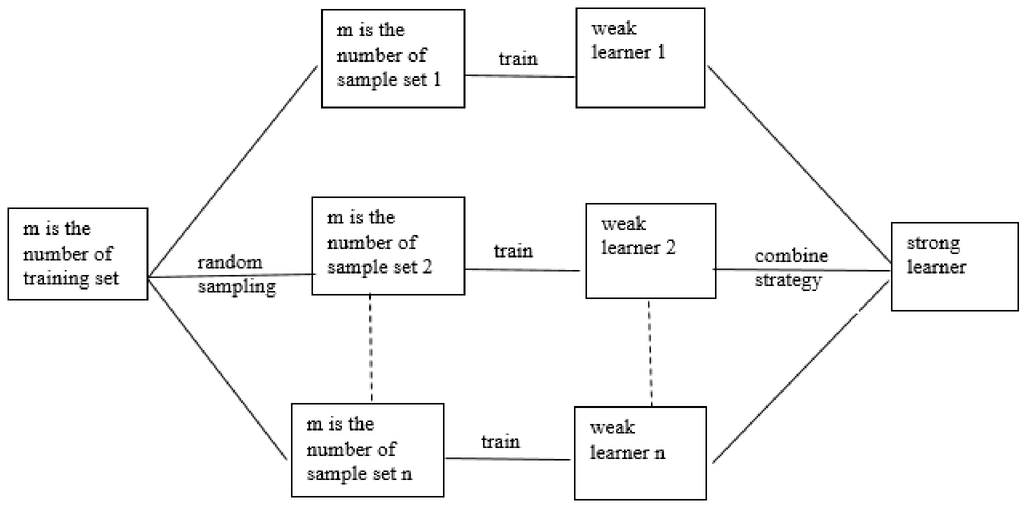

Figure 1. Bagging algorithm working principle (Li, 2022)

In this approach, B different bootstrapped training datasets are generated, then the method is trained on the bth bootstrapped training set to obtain $\hat{\mathbf{f}}^{b}$(x), and finally all predictions are averaged (James, Witten, Hastie, Tibshirani, 2013). This process is called Bagging:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=}^{B} \hat{f}^{*b}(x) \tag{3}$$

2.3. Random Forest

The Random Forest algorithm developed by Leo Breiman in 2001 can be described as an evolutionary version of the Bagging algorithm (Li, 2022).

Using the Random Forest method, Bagging is used along with randomly selected features. Each new training set is drawn back from the original training set using the bootstrap method. Using random feature selection, a tree is then grown on the new training set. Trees that are grown are not pruned (Breiman, 2001).

There is a major difference between the Bagging and Random Forest methods in the selection of m-dimensional variables. During the construction of decision trees, a random sample of m estimators is selected as split candidates from all p estimators, considering each split in a tree (James, Witten, Hastie, Tibshirani, 2013). This random selection of predictors results in a reduction in correlation between trees in the forest, as well as a reduction in variance, which results in a higher accuracy in predictions (Suchetana, Rajagopalan & Silverstein, J., 2017). Also, Random Forests provide some measures of the importance of variables for the prediction of the outcome variable (Gregorutti, Michel, & Saint-Pierre, 2017). The Random Forest algorithm is used only for variable selection in many publications due to this feature.

In the original paper, Breiman (Breiman, 2001) proposed the size of the candidate feature set at each node as **m≈log2(p+1)**. Later, many studies on Random Forest used the default size of the candidate feature set as **m**≈ $\sqrt{p}$ in classification problems and m≈p/3 in regression problems.

Among the difficulties of the algorithm is that the image of the model obtained with the Random Forest algorithm cannot be obtained.

3. APPLICATION AND RESULTS

The happiness score published for 146 countries in the 2022 World Happiness Report and six variables used to explain this score were used in this study. The variables values of GDP levels, life expectancy, generosity, social support, freedom, and corruption in the dataset are not raw data in this study (World Happiness Report, 2022a). The values show the estimated extent to which each variable will contribute to making life assessments higher in each country than in Dystopia (an imaginary country with the most unhappy people in the world) (World Happiness Report, 2022b).

The goal of the study was to figure out how happy or unhappy countries are based on certain rules. We did this by looking at data from the 2022 Happiness Index Report. In order to determine the happiness class of the countries, the 37th country, corresponding to the 1st quartile of 146 countries, was accepted as the boundary. For this reason, countries with a happiness score of 6.3 or above were considered "happy," while other countries were considered "unhappy."

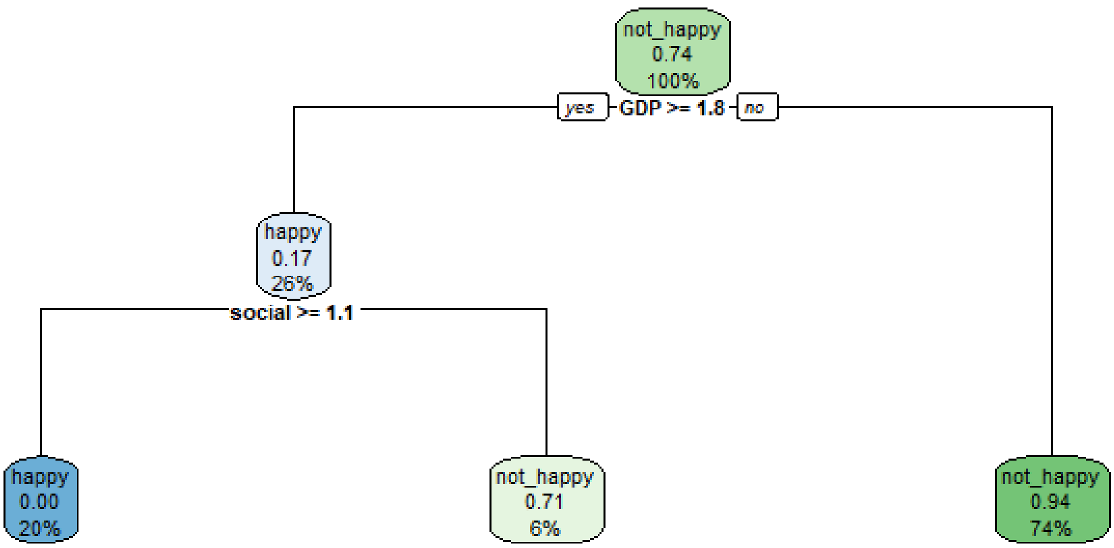

Figure 2. Tree with CART method

As the model must be initially trained and then tested in machine learning, 80% of the data set is randomly allocated as training data and 20% as test data. In the field of machine learning, tree-based CART, Bagging, and Random Forest methods were obtained through the R program, and the outcomes of the methods were analyzed based on the accuracy parameters.

3.1. Model with CART method

In order to construct a classification tree using the CART algorithm, a model is constructed using default values such as minsplit=20 and complexity parameter=0.01.

The results are shown in Figure 2.

The resulting tree shows that the precision value obtained from the training data is 1.00, the recall value is 0.7667, and the F1 value is 0.8679, while the precision value obtained from the test data is 0.8333, the recall value is 0.7143, and the F1 value is 0.7692.

- A country is "happy" if its GDP per capita ≥ 1.8 and social support ≥ 1.
- A country is “not happy” if its GDP per capita ≥ 1.8 and social support < 1.1
- A country is “not happy” if its GDP per capita < 1.8

3.2. Model with Bagging method

500 tree experiments were conducted using the Bagging method. The "mean decrease Gini" values indicate the accuracy of the model when the relevant variable is removed from the model. They are 28.5179 for GDP per capita, 7.4294 for social support, 2.7132 for freedom, 2.3950 for cheating, 1.7719 for corruption, and 1.3569 for generosity. The measurement of "mean decrease accuracy" is a measure of the contribution of the aforementioned variable to the homogeneity of the nodes and their leaves in the resulting Random Forest. In the study, it was found that GDP per capita was 0.1288, social support was 0.0858, freedom was 0.0142, corruption was 0.0008, and generosity was 0.0234. The visuals of these "mean decrease accuracy" and "mean decrease Gini" measures showing the importance of the variables in the model are presented in Figure 3.

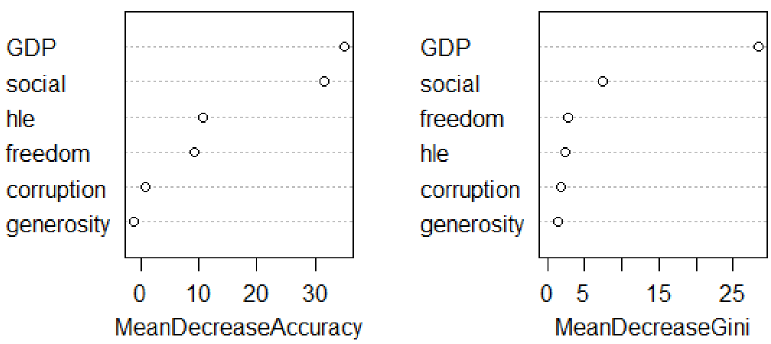

Figure 3. Variable importance plot

According to Figure 3 and the metrics obtained, the most important variables in the model can be stated as GDP per capita and social support, respectively.

The precision value, recall value, and F1 value were obtained in the model obtained with the training data as 1.00. The precision value, recall value, and F1 value of the model as determined by the test data are 0.8571.

3.3. Model with Random Forest method

To determine the most suitable model with this method, trials with a number of randomly sampled variables (mtry) between 2 and 6 as candidates in each compartment were made with the tune process. In the experiments, the number of random variables giving the highest accuracy value was suggested to be 3, as shown in Figure 4.

The "mean decrease Gini" values of the model generated by utilizing the training data and the proposed random three variables are 18.6584 for GDP per capita, 8.7225 for social support, 7.5078 for health, 3.8291 for freedom, 3.4187 for corruption, and 1.8424 for generosity. The "mean decrease accuracy" values of the model generated from the training

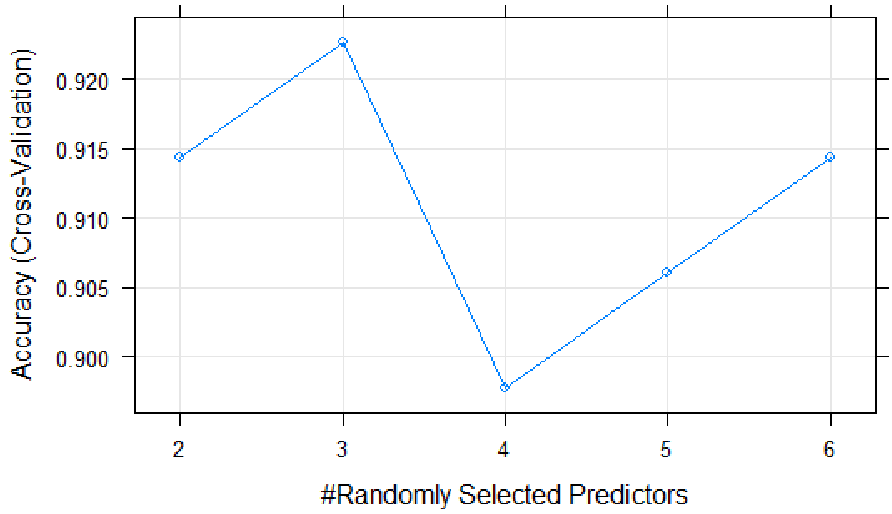

Figure 4. Tune model plot

data with the proposed random three variables are 0.1183 for GDP per capita, 0.0683 for social support, 0.0282 for hle, 0.0181 for freedom, 0.0025 for corruption, and -0.0000 for generosity. The visual representations of the "mean decrease accuracy" and "mean decrease accuracy" measures, which demonstrate the significance of the variables in the model, are depicted in Figure 5.

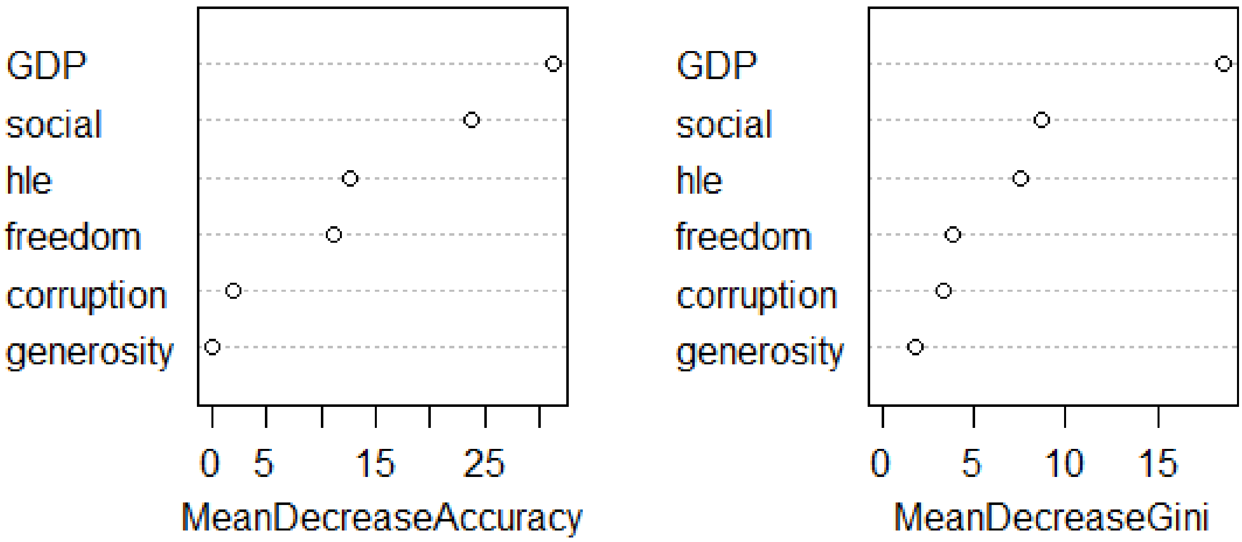

Figure 5. Variable importance plot

According to Figure 5 and the metrics obtained, the most important variables in the model are GDP per capita and social support, respectively. These two variables are the same as the results obtained with the CART and the Bagging method.

In the model constructed using the training data, the precision, recall, and F1 values were all equal to 1.00. The precision value, recall value, and F1 value of the model as determined by the test data are 0.8571.

4. CONCLUSION

The objective of this study was to determine the most significant variables used in the estimation of the World Happiness index of the countries and to determine their significance in the study. Therefore, decision trees, which are commonly used in machine learning, were used. The CART algorithm was employed to generate a single decision tree, while the Bagging and Random Forest algorithms were employed to generate decision trees (decision forests) with ensemble learning.

Upon examination of the model metrics presented in Table 1, it is evident that the test values for precision, recall, and F1 obtained from ensemble learning are superior to those obtained with a single decision tree. This result, as in many previous studies, supports the idea that learning in a community learning gives better performance.

Table 1. *Metrics of tree-based models*

	CART		Bagging		Random Forest	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Precision	1.0000	0.8333	1.0000	0.8571	1.0000	0.8571
Recall	0.7667	0.7143	1.0000	0.8571	1.0000	0.8571
F1	0.8679	0.7692	1.0000	0.8571	1.0000	0.8571

Upon examination of Table 1, it is evident that all tree-based machine learning techniques yield satisfactory outcomes when attempting to classify nations based on their happiness scores. The precision, recall, and F1 values of the Bagging model, which has the advantage of preparing a lower variance model, and the Random Forest method, which has the advantage of reducing the risk of overfitting, are equivalent. According to the literature, this result supports the claim that ensemble-based learning machines have higher performance.

The various reasons for the performance increase in ensemble learning can be stated as follows (Erdem, Uslu & Fırat, 2021):

- Combination of different models reduces overfitting.
- Better data representation and fitting when working with non-linear datasets.
- Reducing class imbalances.
- Increase in calculation performance.

According to all three methods, the two most important variables that determine the happiness classes of countries are GDP and social support. Upon examination of the outcomes obtained through the Random Forest algorithm (Figure 5), it is evident that the variables of healthy life expectancy (HLE) and freedom to make life choices (freedom) hold significant importance subsequent to GDP.

Similar to the findings of this study, Khder, Sayfi, and Fujo (2022) stated that they identified the most significant variables that impact the happiness score through machine learning techniques, namely GDP per capita and health life expectancy.

Jannani, Sael, and Benabbou (Jannani, Sael, & Benabbou, 2021) utilized the 7-year World Happiness Index Report data to generate predictions using diverse machine learning techniques. In this study, the most effective ensemble learning algorithm for prediction was Random Forest, which achieved an R2 value of 0.85. This result supports the conclusion that the Random Forest algorithm is also suitable for the estimation of the happiness index.

As a result of the application and supportive studies in the literature, it can be stated that the classification power of tree-based ensemble learning algorithms for happiness is high. The Bagging and Random Forest algorithms indicate that GDP, social support, and healthy life expectancy are the most important factors in determining happiness class. It is imperative to take into account the foremost priorities of policymakers who aim to enhance the welfare standard of nations.

We believe that dealing with the concept of happiness, which is an important research topic for many fields, will contribute to the literature. Tree-based applications are preferred due to their high accuracy in machine learning and interpretation possibilities. It would be possible to conduct comparative analyses with different happiness indices in future studies.

Peer Review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.

ORCID IDs of the authors / Yazarların ORCID ID'leri
Merve Doğruel 0000-0003-2299-7182
Selin Soner Kara 0000-0002-0894-0772

REFERENCES

Bel, L., Allard, D., Laurent, J. M., Cheddadi, R. & Bar-Hen, A. (2009). CART algorithm for spatial data: Application to environmental and ecological data. *Computational Statistics and Data Analysis*. 53. 3082-3093. https://doi.org/10.1016/j.csda.2008.09.012

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Carlsen, L. (2018). Happiness as a sustainability factor. The World Happiness Index: A posetic-based data analysis. *Sustain Sci. 13.* 549-571. https://doi.org/10.1007/s11625-017-0482-9

Chaudhary, M, Dixit, S. & Sahni, N. (2020). Network learning approaches to study world happiness. A Preprint. https://arxiv.org/pdf/2007.09181.pdf

Dao, T. K. (2017). *Government expenditure and happiness: Direct and indirect effects.* Institute of Social Studies, Netherlands. Retrieved from https://thesis.eur.nl/pub/41656/Dao-Tung-K.-.pdf

Doğruel, M & Fırat, S. Ü. (2021). Veri madenciliği karar ağaçları kullanarak ülkelerin inovasyon değerlerinin tahmini ve doğrusal regresyon modeli ile karşılaştırmalı bir uygulama. *Istanbul Business Research, 50*(2), 465-493. https://www.doi.org/10.26650/ibr.2021.50.015019

Efeoğlu, E. (2022). Kablosuz sinyal gücünü kullanarak iç mekan kullanıcı lokalizasyonu için karar ağaçları algoritmalarının karşılaştırılması. *Acta Infologica.* https://doi.org/10.26650/acin.1076352

Erdem, Z. U., Uslu, B. Ç. & Fırat, S. Ü. (2021). Customer churn prediction analysis in a telecommunication company with machine learning algorithms. *Journal of Industrial Engineering, 32*(3). 496-512.

Farooq, S.A. & Shanmugam, S.K. (2022). A performance analysis of supervised machine learning techniques for COVID-19 and happiness report dataset. In: S. Shakya, V.E. Balas, S. Kamolphiwong, KL. Du (Eds) *Sentimental analysis and deep learning.* Singapore Springer.

Garces, E. J., Adriatico C. & Timbal, L. R. E. (2019). Analysis on the relationships on the global distribution of the World Happiness Index and selected economic development indicators. *Open Access Library Journal, 6,* 1-16. https://doi.org/10.4236/oalib.110545

Grabczewski, K., & Duch, W. (1999, June). *A general purpose separability criterion for classification systems.* Proceedings of the 4th Conference on Neural Networks and Their Application. Zakopane, Poland, 203-208.

Grabczewski, K. (2011), Validated decision trees versus collective decisions. In P. Jedrzejowicz & N. T Nguyen (Eds.) *Computational Collective Intelligence Technologies and Applications Third International Conference* (pp.324-351). Verlag Berlin Heidelberg: Springer.

Gregorutti, B., Michel, B. & Saint-Pierre, P. (2017). Correlation and variable importance in Random Forests. Stat Comput, 27, 659–67. https://doi.org/10.1007/s11222-016-9646-1

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2022). World Happiness Report 2022. New York: Sustainable Development Solutions Network.

Ibnat, Gyalmo, Alom, Abdul Awal and Azim (2021, December). Understanding world happiness using machine learning techniques. International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2). Rajshahi, Bangladesh. https://www.doi.org/10.1109/IC4ME253898.2021.9768407

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R.* New York: Springer.

Jannani, A. Sael, N. & Benabbou, F. (2021, December). *Predicting quality of life using machine learning: Case of World Happiness Index.* 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Alkhobar, Saudi Arabia. https://www.doi.org/10.1109/ISAECT53699.2021.9668429

Khder, M. A., Sayfi, M. A. & Fujo, S. W. (2022). Analysis of World Happiness Report dataset using machine learning approaches. *International Journal of Advances in Soft Computing and its Applications. 14*(1). 14-34. https://doi.org/10.15849/IJASCA.220328.02

Kıroğlu, B. S. & Yıldırım, K. (2022). Mutluluk ve belirleyicileri: Türkiye için bir analiz. *Journal of Emerging Economies and Policy. 7*(2). 5070.

Li, A. (2022). Stock forest model based on Random Forest. In X. Huang & F. Zhang (Eds.) *Economic and business management.* The Netherlands: CRP Press.

Okumuş, F., Ekmekçioğlu, A. & Kara, S. S. (2021). Modelling ships main and auxiliary engine powers with regression-based machine learning algorithms. *Polish Maritime Research, 28*(1). 83-96. https://www.doi.org/10.2478/pomr-2021-0008

Onan, A. (2018). A Clustering Based Classifier Ensemble Approach to Corporate Bankruptcy Prediction. *The Journal of Operations Research, Statistics, Econometrics and Management Information Systems, 6*(2). 365-376.

Öztürk, S. & Suluk, S. (2020). Mutluluk Ekonomisi: G8 ülkeleri açısından ekonomik büyüme ve mutluluk arasındaki ilişkinin incelenmesi. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi.* (37). 226-249.

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion, 27*. 111-125. http://dx.doi.org/10.1016/j.inffus.2015.06.005

Suchetana, B., Rajagopalan, B. & Silverstein, J. (2017). Assessment of wastewater treatment facility compliance with decreasing ammonia discharge limits using a regression tree model. *Science of the Total Environment*, 598, 249- 257. https://doi.org/10.1016/j.scitotenv.2017.03.236

Ulkhaq, M. M. & Adyatama, A. (2021). Clustering countries according to the world happiness report 2019. *Engineering and Applied Science Research, 48*(2). 137-150.

World Happiness Report, 2022a, "Data for table 2.1", https://worldhappiness.report/ed/2022/#appendices-and-data , (29.07.2022)

World Happiness Report, 2022b, "FAQ", https://worldhappiness.report/faq/ , (29.07.2022)

How cite this article

Dogruel, M., Soner Kara, S. (2023). Determining the happiness class of countries with tree-based algorithms in machine learning. *Acta Infologica, 7*(2), 243-252. https://doi.org/10.26650/acin.1251650