doi:10.55385/kastamonujes.1355695 KUJES 9(2):61-66, 2023

Kastamonu University

Journal of Engineering and Sciences

e-ISSN 2667-8209

http://dergipark.gov.tr/kastamonujes

Comparative Performance Analysis of Ensemble Learning Algorithms for Rock Classification

Ebru Efeoğlu

Department of Software Engineering, Faculty of Engineering, Kütahya Dumlupınar University, Kütahya, Türkiye

***Corresponding Author:** ebru.efeoglu@dpu.edu.tr

Received: **September 05, 2023** ◆ Accepted: **November 30, 2023** ◆ Published Online: **December 26, 2023**

Abstract: Knowing the physical and mechanical properties of rocks is important for engineering studies. Because determining the properties and type of rocks affects the safety of engineering structures. Therefore, this study is important in terms of minimizing possible errors in engineering studies. Moreover, Automatic detection of rock types reduces the workload of engineers. In this study, the types of rocks were determined by using some physical and mechanical properties of rocks measured in the laboratory. Rep tree algorithm and ensemble learning algorithms were used in the study. The success of ensemble learning algorithms in classification was compared. As a result, it was understood that ensemble learning algorithms increase success. When the logitboost algorithm was used together with the rep tree algorithm, the Tp rate increased to 0.82. Precision Recall values were 0.80, MCC and AUC were 0.95, kappa was 0.80. In addition, the FP rate decreased to 0.04. The most successful algorithm in rock classification was the Logistboost algorithm. The highest performance metrics were obtained in the classification made with the Logistboost algorithm. In addition, 4 different metric types were calculated to determine the error rates of the algorithms. Logistboost algorithm classified with the lowest error rate.

Keywords: Clasification, Ensemble learning algorithms, Machine learning, Rep tree, Rock

Öz: Kayaların fiziksel ve mekanik özelliklerinin bilinmesi mühendislik çalışmaları açısından önemlidir. Çünkü kayaların özelliklerinin ve türünün belirlenmesi mühendislik yapılarının güvenliğini etkilemektedir. Kaya türlerinin otomatik tespiti mühendislerin iş yükünü azaltır. Bu çalışmada kayaçların laboratuvarda ölçülen bazı fiziksel ve mekanik özellikleri kullanılarak kaya türleri belirlenmiştir. Çalışmada rep ağacı algoritması ve topluluk öğrenme algoritmaları kullanılmıştır. Topluluk öğrenme algoritmalarının sınıflandırmadaki başarısı karşılaştırıldı. Sonuç olarak topluluk öğrenme algoritmalarının başarıyı arttırdığı anlaşıldı. Kaya sınıflandırmasında en başarılı algoritma Logistboost algoritması oldu. Logistboost algoritması ile yapılan sınıflandırmada en yüksek performanslı metrikler elde edildi. Ayrıca algoritmaların hata oranlarını belirlemek için 4 farklı metrik türü hesaplandı. Logistboost algoritması en düşük hata oranına ile sınıflandırmayı gerçekleştirdi.

Anahtar Kelimeler: Sınıflandırma, Topluluk öğrenme algoritmaları, Makine öğrenimi, Rep ağacı, Kaya

1. Introduction

The determination of the properties of the rocks is important in terms of determining the usage area of the rock. The usage area is determined according to the type of rock. For example, basalt is often used as an insulation and durable building material because it is resistant to heat and frost. Basalt as a storage material for high-temperature concentrated solar power plants was investigated by experiments [1]. Fiber reinforced polymeric composite materials were examined [2]. To determine the mechanical properties of basalt, uniaxial compressive strength (UCS) and Young's modulus were evaluated [3]. The geological and mineralogical properties of basalt in Karakalpakstan were investigated [4]. Research has been done on the use of various rocks in tunnel work. Rock tunnel performance under blast loading was investigated by finite element analysis [5]. A three-dimensional finite element analysis of the tunnel was performed under static loading condition and the effect of rock weathering was examined [6]. A simulation model of a rock tunnel with mudstone and sandstone layers was made [7]. The dynamic stability of the subway tunnel in layered weathered sandstone was analyzed [8]. Another name for shale is clay stone and it is generally used in the construction industry. It is used to ensure that the paint is more permanent on the surfaces on which it will be applied. It is used in the manufacture of glass products and materials. In addition, it is of great importance in the extraction and processing of oil, which is very important in the energy sector. The complex conductivity of graphitic schists and sandstones was studied [9]. Granite is used in kitchen counters, sinks in the bathroom, exterior cladding, table and coffee table production, garden stones, decoration decorations. A study was conducted on the use of waste marble and granite dust in structural applications [10]. Experimental studies were carried out on the physical and mechanical changes of hot granite under different cooling processes [11]. Geochemical controls of uranium release from neutral-pH rock drainage produced by the weathering of granite, gneiss and schist [12]. Sandstone is used in construction, paving of roads and pavements, limestone is used in soda production, glass industry and sugar production.

Machine learning methods, which have a wide usage area, were also used in studies related to rocks. Artificial neural network models were used to estimate the unconfined compressive strength of the rock [13]. Supervised machine learning techniques were used to estimate the tunnel boring machine penetration rate [14]. Petrographic classification of sand and sandstone was done [15]. Machine learning methods were used for lithology classification using geophysical data [16]. In this study, it was aimed to automatically detect different rock types using their mechanical and physical properties. and 7 different rock types were identified in the study. Ensemble learning algorithms were used to achieve the best performance.

2. Material and Method

In this study, 7 different rock types collected in Kocaeli were collected [17]. Cylindrical samples with a length of 110–115 mm and a diameter of 54 mm [18,19,20] were prepared from the rock types collected. Ultrasonic pulse velocity, ultrasonic pulse velocity (UPV), uniaxial compressive strength (UCS), resistivity (Ro), chargeability (M) and porosity (n) values of the samples were measured [17]. To measure the UPV value, Proseq acoustic pulse velocity test device was used in accordance with ISRM 1981 and ASTM 1978 standards [17]. The measurements were used to classify rocks. The rocks used and their properties are given in Table 1.

Table 1. Rocks and their properties

Rocks	The properties of the rocks
Lime stone	It is a type of sedimentary rock with at least 90% calcium carbonate ($CaCO_3$) in its chemical composition and at least 90% calcite mineral in its mineralogical composition.
Arkose	It is a type of sandstone. They derive from granitoid igneous rocks. It contains feldspar and mica fractures.
Sandstone	It is a type of sedimentary rock. Contains feldspar and quartz.
Basalt	It is a type of igneous rock. It contains high amounts of MgO and CaO and low amounts of $Na_2O + K_2O$.
Granite	Granite is a hard, crystalline mineralized igneous rock. It contains silicate, quartz and feldspar minerals.
Amphibolite	It is a metamorphic rock containing amphibole hornblende and actinolite.
Schist	It is a metamorphic rock

Rep Tree

REP Tree algorithm is a fast algorithm and works with regression tree logic. The information gain criterion is used to construct the regression tree. It creates multiple trees in iterations and then chooses the best one from the trees it creates. The pruning method is used to minimize the error.

Bagging

Bagging algorithm is an ensemble learning method for creating a classifier ensemble by combining basic learning algorithms trained on different samples of the training set [21]. The Bagging algorithm is based on the principle that each basic learning algorithm that makes up the community is trained on different training sets. Thus, diversity increases. It is aimed to create different training sets from the data set. Simple random substitution sampling method is generally used to create the training set. Then, the outputs of the classification methods trained with the training sets are combined through majority voting.

Adaboost

AdaBoost algorithm is one of the most basic Boosting methods. This algorithm focuses more on samples that are difficult to classify. The purpose of the algorithm is to increase the classification success [22]. Weight values are changed with iterations. In each iteration, the weight values of the correctly classified samples are decreased and the weight values of the incorrectly classified samples are increased. This allows more iterations to be allocated to data samples that are difficult to classify.

Logitboost

AdaBoost algorithm can be affected by noises and overfitting problem may occur. It therefore recommends using LogitBoost for noisy data [23]. Training errors can be reduced by using the Loogitboost algorithm.

3. Result

In this study, laboratory test results were used to determine the physical and mechanical properties of 7 different rock types. Rep tree algorithm and ensemble learning algorithms from machine learning techniques were used to detect the types of rocks. The tree structure created by the software for the rep tree algorithm is given in Figure 1. Accordingly, the root node was chosen as M and the first branching was done. The first branching was done according to whether the M value was less than 14.25 or greater than 14.25. If the M value is greater than or equal to 14.25, the Ro value was checked and the rock type was determined. In the other part of the tree, if the n value is greater than or equal to 0.08, the upv value was checked. If the UPV value was less than 4035 and Schist was larger, it was classified as Amphibolite. In other branches, n and UCS values were checked and other samples were classified.

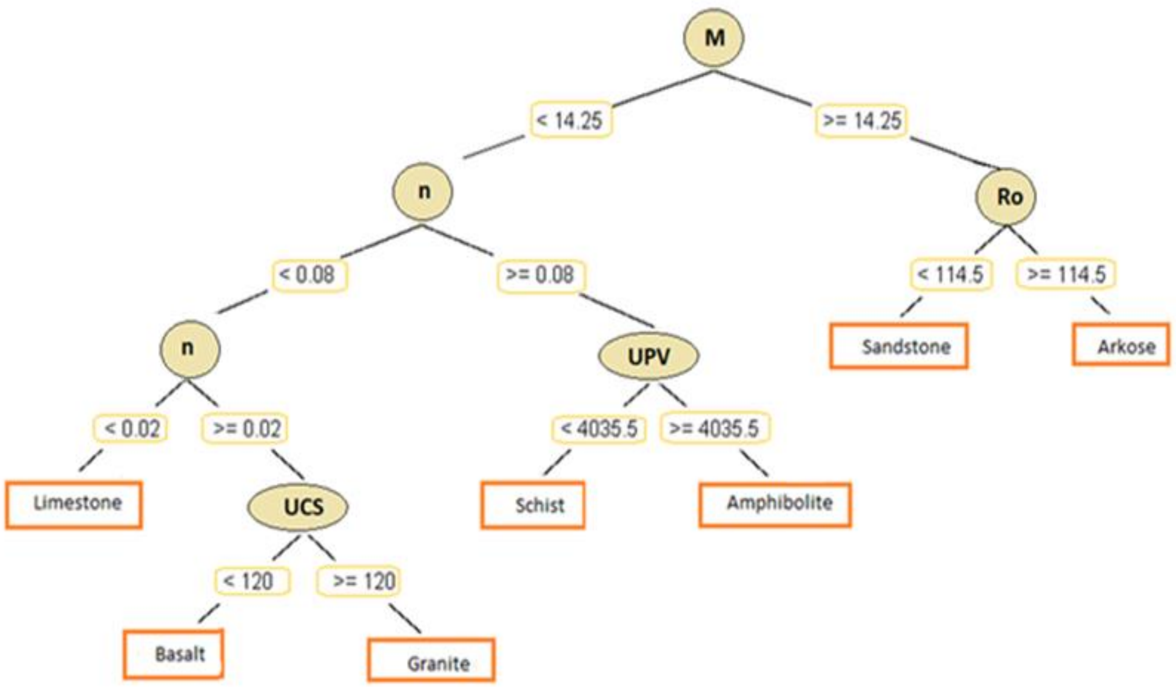

Figure 1. Rep Tree

To improve the performance of the rep treee algorithm, 3 different ensemble learning algorithms were used and their results were compared. For comparison, firstly, the individual confusion matrices of the algorithms are given in Table 2,3,4,5, respectively. It was seen in Confusion matrices that the number of correctly predicted instances increased when ensemble learning algorithms were used with the Rep tree algorithm. The highest increase in the number of samples was seen when the Logitboost algorithm was used.

Table 2. Confusion matrix of Rep tree

		Predicted						
		Limestone	Arkose	Sandstone	Basalt	Granite	Amphibolite	Schist
Actual	Limestone	**10**	0	0	2	0	0	0
	Arkose	0	**10**	8	0	0	0	0
	Sandstone	0	10	**10**	0	0	0	0
	Basalt	0	0	0	**9**	1	0	0
	Granite	0	0	0	1	**9**	0	0
	Amphibolite	0	0	0	0	0	**13**	1
	Schist	0	0	0	0	0	0	**12**

Table 3. Confusion matrix of Bagging+Rep Tree

		Predicted						
		Limestone	Arkose	Sandstone	Basalt	Granite	Amphibolite	Schist
Actual	Limestone	**10**	0	0	2	0	0	0
	Arkose	0	**6**	12	0	0	0	0
	Sandstone	0	6	**14**	0	0	0	0
	Basalt	0	0	0	**10**	0	0	0
	Granite	0	0	0	0	**10**	0	0
	Amphibolite	0	0	0	0	0	**13**	1
	Schist	0	0	0	0	0	0	**12**

Table 4. Confusion matrix of Adaboost+rep tree

		Predicted						
		Limestone	Arkose	Sandstone	Basalt	Granite	Amphibolite	Schist
Actual	Limestone	**10**	0	0	2	0	0	0
	Arkose	0	**11**	7	0	0	0	0
	Sandstone	0	8	**12**	0	0	0	0
	Basalt	0	0	0	**9**	1	0	0
	Granite	0	0	0	1	**9**	0	0
	Amphibolite	0	0	0	0	0	**13**	1
	Schist							**12**

Table 5. Confusion matrix of Logitboost+rep tree

		Predicted						
		Limestone	Arkose	Sandstone	Basalt	Granite	Amphibolite	Schist
Actual	Limestone	**12**	0	0	0	0	0	0
	Arkose	0	**8**	10	0	0	0	0
	Sandstone	0	9	**11**	0	0	0	0
	Basalt	0	0	0	**10**	0	0	0
	Granite	0	0	0	0	**10**	0	0
	Amphibolite	0	0	0	0	0	**14**	0
	Schist	0	0	0	0	0	0	**12**

Other performance metrics of the algorithms were calculated and given in Table 6. The farther the metric values are from 0, the more successful the algorithm is. These values can be at most 1. The lowest metric values were obtained when Ensemble learning algorithms were not used. Bagging, Adaboost and Logit boost algorithms increased all performance metrics. The highest TP Rate was 0.82, the highest Precision, recall, F-Score was 0.80, and the highest AUC was 0.95 when Logitboost+Reptree algorithms were used. The accuracy rates obtained by the algorithms in classification are given in Figure 2. According to the figure, the highest accuracy value is 80%. This value was obtained when the loogit boost algorithm and the Reeptree algorithm were used together.

Table 6. Performance metrics of algorithms

	TP Rate	FP Rate	Precision	Recall	F-Score	MCC	AUC	Kappa
Rep Tree	0.76	0.05	0.76	0.76	0.76	0.71	0.91	0.71
Bagging+Rep Tree	0.78	0.05	0.78	0.78	0.77	0.73	0.94	0.74
Adaboost+Reptree	0.79	0.04	0.79	0.79	0.79	0.74	0.95	0.75
Logitboost+Reptree	0.82	0.04	0.80	0.80	0.75	0.95	0.95	0.76

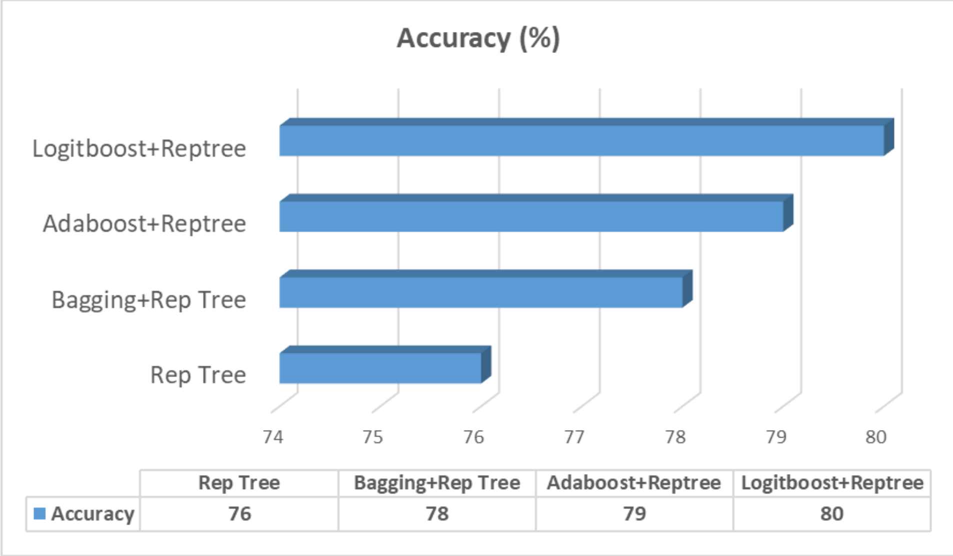

Figure 2. The accuracy of the algorithms

Another metric used in performance evaluation is the error metric. There are many error metrics developed for this. In the study, the error rates of the algorithms were calculated using 4 different metrics and are given in Table 7. It was seen in the table that the lowest error value was obtained with the Loogit boost algorithm.

Table 7. Error metrics of algorithms

	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error
Rep Tree	0.0682	0.2186	28.1285	62.7637
Bagging+Rep Tree	0.0766	0.2053	31.5813	58.9681
Adaboost+Reptree	0.0625	0.2253	25.7863	64.707
Logitboost+Reptree	0.0527	0.183	21.7339	52.5597

4. Discussion and Conclusion

In this study, 7 different rock types were classified using machine learning techniques. Rep tree algorithm, a decision tree algorithm, was used in classification. For classification, a data set containing ultrasonic pulse velocity, ultrasonic pulse velocity, uniaxial compressive strength, resistivity, chargeability and porosity values, which are used to determine the physical and mechanical properties of rocks, was used. As a result of the classification, various performance metrics were calculated to determine the performance of the algorithm. TP rate, Precision Recall and F-Score values were found to be 0.76, MCC and Kappa values were 0.71, AUC value was 0.91 and FP Rate value was 0.05. Three different Ensemble Learning Algorithms were then used to improve the performance of the algorithm. As a result, it was seen that the three algorithms used increased the classification performance. logitboost algorithm was the best performing algorithm among Ensemble Learning Algorithms. When the logit boost algorithm was used together with the rep tree algorithm, the Tp rate increased to 0.82. Precision Recall values were 0.80, MCC and AUC were 0.95, kappa was 0.80. In addition, the FP rate decreased to 0.04. For this reason, it is recommended to use the rep teree algorithm and logitboost algorithm together in rock classification.

Conflict of Interest

The authors declare that they have no competing interests.

Author Contribution

We declare that all Authors equally contribute.

Funding

The author declared that this study has received no financial support.

5. References

[1] Nahhas, T., Py, X., & Sadiki, N. (2019). Experimental investigation of basalt rocks as storage material for high-temperature concentrated solar power plants. Renewable and Sustainable Energy Reviews, 110, 226-235.
[2] Fegade, V., Ramachandran, M., Madhu, S., Vimala, C., Malar, R. K., & Rajeshwari, R. (2022, May). A review on basalt fibre reinforced polymeric composite materials. In AIP Conference Proceedings 2393(1). AIP Publishing.
[3] Aldeeky, H., Al Hattamleh, O., & Rababah, S. (2020). Assessing the uniaxial compressive strength and tangent Young's modulus of basalt rock using the Leeb rebound hardness test. Materiales de Construcción, 70(340), e230-e230.
[4] Mustapaevich, D. K. (2021). Geological-Geochemical and Mineralogical Properties of Basalt Rocks of Karakalpakstan. International Journal on Integrated Education, 4(10), 205-208.
[5] Sadique, M. R., Zaid, M., & Alam, M. M. (2022). Rock tunnel performance under blast loading through finite element analysis. Geotechnical and Geological Engineering, 40(1), 35-56.
[6] Zaid, M. (2021). Three-dimensional finite element analysis of urban rock tunnel under static loading condition: effect of the rock weathering. Geomechanics and Engineering, 25(2), 99-109.
[7] Zhao, D., He, Q., Ji, Q., Wang, F., Tu, H., & Shen, Z. (2023). Similar model test of a mudstone-interbedded–sandstone-bedding rock tunnel. Tunnelling and Underground Space Technology, 140, 105299.
[8] Shoeb, M., Khan, S. A., Alam, T., Ali, M. A., Gupta, N. K., Ansari, M. M., ... & Dobrota, D. (2023). Dynamic stability analysis of metro tunnel in layered weathered sandstone. Ain Shams Engineering Journal, 102258.
[9] Abdulsamad, F., Revil, A., Ghorbani, A., Toy, V., Kirilova, M., Coperey, A., ... & Ravanel, L. (2019). Complex conductivity of graphitic schists and sandstones. Journal of Geophysical Research: Solid Earth, 124(8), 8223-8249.
[10] Nayak, S. K., Satapathy, A., & Mantry, S. (2022). Use of waste marble and granite dust in structural applications: A review. Journal of Building Engineering, 46, 103742.
[11] Kang, F., Jia, T., Li, Y., Deng, J., & Huang, X. (2021). Experimental study on the physical and mechanical variations of hot granite under different cooling treatments. Renewable Energy, 179, 1316-1328.
[12] Skierszkan, E. K., Dockrey, J. W., Mayer, K. U., Bondici, V. F., McBeth, J. M., & Beckie, R. D. (2020). Geochemical Controls on Uranium Release from Neutral-pH Rock Drainage Produced by Weathering of Granite, Gneiss, and Schist. Minerals, 10(12), 1104.
[13] Barham, W. S., Rabab'ah, S. R., Aldeeky, H. H., & Al Hattamleh, O. H. (2020). Mechanical and physical based artificial neural network models for the prediction of the unconfined compressive strength of rock. Geotechnical and Geological Engineering, 38, 4779-4792.
[14] Xu, H., Zhou, J., G. Asteris, P., Jahed Armaghani, D., & Tahir, M. M. (2019). Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. Applied sciences, 9(18), 3715.
[15] Garzanti, E. (2019). Petrographic classification of sand and sandstone. Earth-science reviews, 192, 545-563.
[16] Bressan, T. S., de Souza, M. K., Girelli, T. J., & Junior, F. C. (2020). Evaluation of machine learning methods for lithology classification using geophysical data. Computers & Geosciences, 139, 104475.
[17] Sertçelik, İ., Kurtuluş, C., Sertçelik, F., Pekşen, E., & Aşçı, M. (2018). Investigation into relations between physical and electrical properties of rocks and concretes. Journal of Geophysics and Engineering, 15(1), 142-152.
[18] ASTM Standard 1984Standard test method for unconfined compressive strength of intact corespecimens soil and rock building stones Annual Book of ASTM Standards 4.08
[19] ASTM Standard 2001Standard Practice for Preparing Rock Core Specimens and Determining Dimensional and Shape Tolerances (Philadelphia, PA: American Society for Testing and Materials) D4543
[20] ISRM 2007 The Complete ISRM Suggested Methods for Rock Characterization Testing and Monitoring:1974–2006 ed R Ulusay and J A Hudson (Ankara: Kozan Ofset Press)
[21] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
[22] Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In Icml 96, 148-156.
[23] Friedman, J., Hastie T., & Tibshirani, R. (2000) Additive Logistic Regression: a Statistical View of Boosting, Annals of Statistics, 28, 337-407.