Journal of Information Systems and Management Research
Bilişim Sistemleri ve Yönetim Araştırmaları Dergisi

http://dergipark.gov.tr/jismar

Araştırma Makalesi / Research Article

Gelişmekte Olan Ülkelerde Matematik Başarısını Etkileyen Faktörlerin Araştırılmasında Makine Öğrenme Tekniklerinin Kullanılması: Türkiye, Meksika, Tayland Ve Bulgaristan Örneği

Mahmut ÇAVUR[a], Tuba ARPA

[a,*] *Kadir Has Üniversitesi, Lisansüstü Eğitim Fakültesi, Yönetim Bilişim Sistemleri.Bölümü, İSTANBUL, 34083, TÜRKİYE*

MAKALE BİLGİSİ

Alınma: 13.07.2024
Kabul: 21.12.2024

Anahtar Kelimeler:
PISA, makine öğrenmesi, öğrenci başarısı, matematik başarısı, algoritmaları karşılaştırmak

Sorumlu Yazar
tubaarpa@gmaiil.com

ÖZET

Bu çalışmada, PISA 2018 verileri kullanılarak, Türkiye, Bulgaristan, Meksika ve Tayland'daki öğrencilerin başarılarını etkileyen faktörlerin, öğrenci üzerindeki etkisinin tespitinde çeşitli makine öğrenimi modellerinin etkinliği karşılaştırılmıştır. Çalışmada regresyon için; doğrusal regresyon, destek vektör makinesi, karar ağacı ve rastgele orman, sınıflandırma için; lojistik regresyon, destek vektör makinesi, karar ağacı ve rastgele orman modelleri kullanılmıştır. Ayrıca, XGBoost matematik başarısının temel belirleyicileri tanımlanmış ve K-Means kümeleme ile eksik verileri doldurulmuştur. Sonuçlara göre, tüm ülkeler için, öğrencilerin ekonomik ve sosyokültürel durumları, evdeki çalışma materyalleri, sorumluluk duyguları ve ailelerinin ilgisi temel katkı faktörlerini oluşturmaktadır. Model başarısı açısından, rastgele orman modeli hem regresyon hem de sınıflandırmada diğer modellere göre daha başarılı olmuş, rastgele orman regresyonu en yüksek R-kare değerlerini (%71-%84) elde etmiştir, doğrusal regresyon ise en düşük değerleri (%22-%43) vermiştir. Buna ek olarak, sınıflandırma algoritmaları ikili ve üçlü sınıflandırma açısından da analiz edilmiş, ikili sınıflandırmanın üçlü sınıflandırmadan daha başarılı olduğu gözlemlenmiştir. Rastgele orman algoritmasının doğruluk skorları ülkeler arasında %73 ile %83 arasında değişmiştir. Çalışmanın bulguları, öğrencinin matematik başarısına etki eden faktörleri tahmin etmek için en uygun algoritmaların seçiminde, karar vericiler için değerli içgörüler sunmakta ve eğitim sonuçlarını iyileştirmeleri için karar vericilere yardımcı olmaktadır.

DOI: 10.59940/jismar.1514958

A Comparative Analysis of Machine Learning Techniques to Explore Factors Affecting Mathematics Success in Developing Countries: Türkiye, Mexico, Thailand, And Bulgaria Case Studies

ARTICLE INFO

Received: 13.07.2024
Accepted: 21.12.2024

Keywords:
PISA, machine learning, students' achievement, mathematics achievement, comparing algorithms

Corresponding Authors
tubaarpa@gmail.com

ABSTRACT

This study explores factors influencing mathematics achievement in Türkiye, Bulgaria, Mexico, and Thailand using PISA 2018 data and machine learning models, comparing their performance. Both classification and regression models were utilized: linear regression, support vector machine, decision tree, and random forest for regression; logistic regression, support vector, decision tree, and random forest for classification. Additionally, XGBoost identified key predictors of math achievement, and K-Means filled missing data. According to results, key contributing factors across all countries included students' economic, social, and cultural status, study materials at home, sense of ownership, and family welfare. Regarding model success, random forests outperformed other models in both regression and classification, with Random Forest Regression achieving the highest R-square values (71%-84%) while linear regression has the lowest (22%-43%). In addition, the classification algorithms were analyzed in terms of binary and ternary classification, binary classification proved more successful than ternary, with RF accuracy scores ranging from 73% to 83%

across countries. The study's findings offer valuable insights for selecting optimal algorithms for predicting math achievement, aiding decision-makers in enhancing educational outcomes.

DOI: 10.59940/jismar.1514958

1. *INTRODUCTION (GİRİŞ)*

Mathematics is of vital importance for the development and progress of a society. It is considered one of the most critical areas of educational systems because it enables individuals to develop their cognitive abilities and plays a vital role in the basis of advanced technological and scientific research. In addition to developing students' analytical thinking, problem-solving, and critical reasoning skills, mathematics skills prepare students to succeed in today's complex world [1]. In this context, students' mathematics achievement provides essential information about the quality of a country's education system. The relationship between the level of development of countries and the education system has long attracted the attention of researchers. In literature, it is emphasized that education strongly impacts economic growth in two aspects. Firstly, human capital, which refers to people's mental and physical strength, is an input in the production function. Secondly, human capital is essential in research that produces technology and knowledge [2]. Also, the role of mathematics education within the education system is crucial because mathematics plays a vital role in our daily lives [3]. In this context, increased spending on education can have profound consequences for developing countries because education in general, and mathematics education in particular, can be an increasingly effective instrument for boosting a country's GDP growth [4]. Analyzing the current educational situation well is important to provide all these benefits. Therefore, countries must understand society's education level and develop policies accordingly. Various institutions and organizations must make objective assessments, measure, report, and present the educational achievement of countries. PISA is essential as it is an internationally recognized assessment tool due to its role and importance in mathematics measurement. PISA results offer thorough and comparable information on students' math proficiency. Due to the big data, high number of observations, and variables presented by PISA, it is almost necessary to apply machine learning methods for prediction and inferential statistical models with the data obtained from PISA. In this context, machine learning models are an essential tool in evaluating PISA data. In support of this, it is seen in the literature that the techniques applied in developing prediction models related to education have increased towards machine learning models [5].

Consistent with the above, the main motivation of this study is to use machine learning modeling to analyze the factors that influence students' math achievement across countries of similar economic size and level of development, including Turkey, Thailand, Bulgaria, and Mexico. The relevant machine learning methods include eight models, four regression, and four classification algorithms. These models are Multiple Linear Regression, Support Vectorial Regression, Decision Tree, Random Forest, Logistic, Support Vectorial Classification, Decision Tree, and Random Forest. The relationship between students' mathematics achievement and potential influencing factors is examined through regression analyses. Classification analyses will be used to group students according to specific achievement levels and evaluate factors' impact on classification performance. In addition, the success of each model will be compared through various metrics to guide researchers interested in studying the subject.

2. LITERATURE REVIEW *(Literatür Taraması)*

Students' achievement in mathematics is considered a strong indicator of academic success in the years to come [6] and is correlated with countries' levels of development and GDP [4]. Understanding mathematics achievement at the national level is a complex and difficult process due to big data, so there are various measurement methods. One of these methods is the Program for International Student Assessment (PISA), carried out internationally. It is an exam conducted by The Organization for Economic Co-operation and Development (OECD) to evaluate the reading, comprehension, science, and mathematics skills and knowledge levels of students in the age group of 15 who have completed their compulsory education. PISA, an international exam, focuses not only on measuring course success but also on how well they can make sense of this information in school and out-of-school environments and how well they can apply it in different situations. In addition to measuring student achievement, increasing the functionality of the education system, determining the effects of education policies on students, and increasing the quality of education are among the objectives of PISA [7],[8],[9],[10]. Türkiye has been involved in PISA studies since 2003. Türkiye participated in computer-based applications in 2015 and 2018 [9]. Because decision-makers in Türkiye believe that they can use

PISA data and analysis as an essential resource for developing education policies and improving education systems [11]. Machine learning is frequently used to analyze complex and big data. Machine learning is a branch of artificial intelligence which enables computers to learn from data, examples and training. It involves learning to identify significant patterns in large datasets. The availability of large amounts of data has made it easier to train machine learning systems, while advances in computer processing power have increased the capacity of these systems [12]. One of the main advantages of machine learning algorithms is their ability to make evidence-based decisions by analyzing large and complex datasets. This has improved decision-making in various fields, including healthcare, manufacturing, education, finance, policing, and marketing [13]. Furthermore, selecting the correct machine-learning algorithm for a given task is critical. The nature of the data and mission can affect an algorithm's performance, and different algorithms have varied strengths and weaknesses. To choose the best algorithm for a specific task, empirical comparisons and assessments of several algorithms are crucial [14]. Machine learning tasks usually fall into three broad categories: Supervised, Unsupervised, Semi-Supervised, and Reinforcement learning and it is seen that supervised algorithms are the most widely used algorithms in the field of education [5]. When we examine mathematical literacy based on PISA data, it is seen that most of them use one or two different machine-learning models. On the other hand, Lezhnina and Kismihók [15] used only the random forest algorithm in the study in which they wanted to combine statistical and machine learning methods. Also, Güre, Kayri, and Erdoğan [16] compared only neural networks and random forest algorithms in their study for comparison purposes. Finally, studies that use many machine learning algorithms mostly use classification algorithms. For example, although Saarela et al. [17] used five different machine learning methods in their study, all are classification algorithms. However, in this study, both regression and classification algorithms were used, thus comparing the performance of the results when the target variable is continuous and categorical. This is thought to be an essential contribution to literature and will guide future studies on what the target variable should be.

3. METHODOLOGY *(Yöntem)*

The PISA dataset published every three years by the OECD is used in this study. It is based on the most recent publicly available data from PISA 2018. The PISA 2018 data cover 612004 students from 21903 schools in 79 countries and economies. Student literacy in reading, science and mathematics is measured in the dataset. Four countries were included after filtering the country variable in the dataset according to the purpose of the study. These countries were selected from Europe, Asia and America, which are close to Turkey in economic size. The aim was to have a comparison of countries with similar economic measures (GDP). Turkey, Bulgaria, Thailand and Mexico were the countries of choice for the study. In this context, the data of 28117 students from the 4 countries have been analysed in the context of the research.

3.1 DATA *(Veri)*

Table 1. *Variables Description (Değişken Açıklamaları)*

Variables	Description	Data Type	Scale Type
CNT	Country	String	Nominal
ST004D01T	Gender	String	Nominal
ESCS	Index of economic social and cultural status	Numeric	Interval
WEALTH	Family wealth possession	Numeric	Interval
HOMEPOS	Index of all household and possession items	Numeric	Interval
CULTPOSS	Cultural possessions	Numeric	Interval
HEDRES	Home educational resources	Numeric	Interval
ST011Q02TA	Having personal room	String	Nominal
MISCED	Level of mother education	Numeric	Ordinal
FISCED	Level of father education	Numeric	Ordinal
MMINS	Learning times in Math (per minutes at week)	Numeric	Ratio
TMINS	Learning times (per minutes at week)	Numeric	Ratio
PERCOOP	Index of student co-operation	Numeric	Interval
PERCOMP	Index of student competition	Numeric	Interval
BELONG	Index of sense of belonging	Numeric	Interval
EMOSUPS	Index of parents' emotional support	Numeric	Interval
PERFEED	Teacher feedback	Numeric	Interval
PVMATH	Plausible value in Math	Numeric	Interval

In the PISA dataset, there are ten mathematically plausible scores provided. These scores are not typical individual student scores; instead, they represent a range of possible abilities a student might possess based on their responses to test items. Utilizing item response theory (IRT), ten plausible values (PVs) are generated by sampling from the posterior probability distribution of the ability estimates [18]. One of these PVs can be selected randomly, or a new score can be derived by calculating the average of all ten values. In this research, the target variable was created by taking the mean of these ten distinct PV values.

Apart from the gender variable, the remaining variables in this dataset represent indices generated by PISA. Most of the metrics in PISA can be seen as indices that aggregate responses from students, parents, teachers, or school officials (typically principals) to a set of related questions. These questions were chosen from a broader selection, based on theoretical frameworks and prior studies. To assess the effectiveness of the machine learning models, the dataset in this study was split into training and testing sets. Specifically, 80% of the data was allocated for training, while the remaining 20% was reserved for testing.

3.2 Regression and Classification Algorithms *(Regresyon ve Sınıflandırma Algoritmaları)*

Machine learning, a subset of artificial intelligence in computer science, enables systems to learn and improve from experience using data. Rooted in statistical learning theories, these algorithms apply statistical and computational techniques to detect patterns in data and forecast future trends. Depending on the training approach and whether outputs are provided during training, machine learning can be divided into ten distinct categories. The categories include neural networks, dimensionality reduction techniques, supervised, unsupervised, semi-supervised, ensemble, reinforcement, instance-based learning, evolutionary, and hybrid approaches. [19]. This study used a total of eight different machine learning models. Four of these models consisted of regression and four consisted of classifying algorithms. The classification algorithms are logistic regression, logistic regression and logistic regression. The classification algorithms are Logistic, Vectorial Classification, Decision Tree, and Random Forest. The goal is the comparison of regression and classification techniques in the prediction of the target variable. The scikit-learn and stats-models libraries in Python are used to implement these algorithms. All the algorithms used in the research are briefly explained in the following sections.

3.2.1 Regression Algorithms *(Regresyon Algoritmaları)*

As a statistical technique, regression analysis is widely used to understand the relationship between dependent and independent variables. This analysis enables the investigator to comprehend and forecast how the value of the dependent variable will change with a change in any of the independent variables. [20].

Linear Regression (Doğrusal Regresyon)

Linear regression is a common statistical approach utilized in machine learning to model the connection between a dependent variable (target) and one or several independent variables. This relationship is represented through a linear equation. When there is only a single predictor, the method is known as simple linear regression. However, if there are multiple predictors involved, the model is referred to as multiple linear regression [21].

Support Vector (Destek Vektörü)

Support Vector Regression (SVR) is an adaptation of the Support Vector Machine (SVM) designed for regression tasks, aiming to fit a continuous function to the given data. SVR retains many of the strengths of SVM classification, including its capability to manage high-dimensional datasets and capture intricate relationships among variables. The use of a kernel function enables the algorithm to map nonlinear relationships into a higher-dimensional space, where linear patterns can be more easily detected, thus allowing SVR to effectively handle nonlinear dependencies between variables. This allows the SVR to capture complex patterns and predict accurately in the presence of noise or imperfect data [22].

Decision Tree Regression (Karar Ağacı Regresyonu)

Decision tree regression is essentially an adapted version of decision tree classification for approximating real-valued functions such as proportions or continuous variables. This method proceeds by subdividing the data through a process of repeated binary splitting. Decision tree regression creates a tree-like structure for modeling real-valued functions and at each step selects the optimal split that minimizes the sum of squared deviations. This process continues until the minimum node size of the tree is reached. The resulting tree provides a predictive

model for continuous variables [23].

Random Forest Regression (Rastgele Orman Regresyonu)

Random forest regression is an ensemble method that combines multiple decision trees to build a strong predictive model suited for regression problems. While initially developed for classification purposes, it has been adapted and extended for regression analysis. In random forest regression, multiple decision trees are constructed and combined to enhance the overall prediction accuracy. An arbitrary training data set is used to train each tree, using an arbitrary feature set. During training, each tree independently predicts on the basis of the input [24].

3.2.2 Classification Algorithms (Sınıflandırma Algoritmaları)

Classification algorithms are techniques designed to analyze data that has already been categorized. In this context, classification problems arise when the outcome is restricted to one of several predefined categories, such as "Yes/No" or "True/False." Based on the number of potential output classes, the problem is classified as either a binary classification (with two classes) or a multiclass classification (with more than two classes) [19].

When the outcome is binary, such as determining whether a student has a personal room, the model is referred to as a binary logistic model. If the logistic regression model includes only a single predictor variable, it is known as simple logistic regression. However, if the model includes multiple predictor variables, which can be either categorical or continuous, it is termed as multiple or multivariate logistic regression [25].

Support Vector Classification: Support Vector Classification (SVC) is an algorithm in machine learning designed to address classification tasks. This approach utilizes a learned decision boundary, known as a hyperplane, to separate and classify data points effectively. SVC is a classification variant of the Support Vector Machine (SVM) and employs the same foundational principles as SVM. It uses the same support vector concept as SVM to determine a decision limit for classifying data points into different classes. The main advantage of the SVC is that it is able to make the data separable in a linear way in high-dimensional spaces. The data points are classified through the creation of complex decision boundaries [26].

Decision Tree: A decision tree is a hierarchical model resembling a flowchart, where rectangles denote internal decision nodes and ovals indicate leaf nodes. This algorithm is widely used because it is simpler to implement and more intuitive than many other classification methods [27]. Decision tree classifiers often provide comparable or even superior accuracy compared to alternative classification techniques. Depending on the dataset size, available computational resources, and the algorithm's scalability, decision trees can be executed in a sequential or parallel manner [28].

Random Forest: Random forest is an ensemble technique that merges multiple machine learning and classification algorithms. It aggregates the predictions from a collection of decision trees, with each tree casting a single vote for the most likely class. The combined results of these votes determine the final classification. Random forests typically exhibit high accuracy, are resilient to outliers and noise, and avoid overfitting issues [29].

3.3 Feature Selection *(Özellik Seçimi)*

Selecting the right features or variables is a critical phase in constructing a machine learning model. The inclusion or exclusion of specific variables can significantly alter the overall performance of the model. In this study, multiple approaches were applied for feature selection. The first approach involved leveraging findings from existing literature. As described earlier, the variables chosen in this research either directly align with those mentioned in the literature or correspond to the 2018 equivalents of previously studied variables. The second approach focused on assessing variable importance using the XGBoost classification algorithm.

XGBoost (eXtreme Gradient Boosting) is a widely used algorithm known for its effectiveness in feature selection. One of the primary strengths of XGBoost is its ability to highlight key features in a dataset. By utilizing metrics like feature importance ranking and feature contribution, it helps in identifying the most influential variables. This approach aids in removing irrelevant or less impactful features, ultimately enhancing the model's predictive capability.

3.4 Model Evaluation and Metrics *(Model Değerlendirme ve Metrikler)*

Evaluation metrics are essential standards used to assess the effectiveness of classification algorithms in an objective manner. In this research, the key metrics utilized to evaluate classification performance include

the Confusion Matrix, Accuracy, Precision, Recall (Sensitivity), F1 Score, and ROC Curve.

These evaluation metrics are applied to compare the performance of various classification algorithms, helping to identify the model that achieves the highest performance. They provide a standardized approach for determining which algorithm excels in specific metrics, offering a fair basis for model selection.

Accuracy is defined as the proportion of correctly predicted instances out of the total number of instances. The calculation for this metric follows the formula shown below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision refers to the ratio of correctly identified positive samples among all samples predicted as positive. The formula below is used to determine the precision value.

$$Precision = \frac{TP}{TP + FP}$$

Recall indicates the rate at which true positive samples get detected. This measure is used according to this formula.

$$Recall = \frac{TP}{TP + FN}$$

The F1 value gives the harmonious mean of sensitivity and precision. It is a metric that is often used as a balanced metric for evaluation purposes. The formula that is used to calculate this metric is given below.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

The ROC curve illustrates the connection between the true positive rate (TPR) and the false positive rate (FPR) by varying the classification thresholds. The area under the curve (AUC) quantifies the area beneath the ROC curve and serves as an indicator of the classifier's overall performance.

Also, in this study, we used Mean Absolute Error (MAE) and R-squared, two important evaluation metrics for regression analysis. These two metrics are widely utilized statistical tools for assessing the performance of regression models. Mean Absolute Error (MAE) calculates the mean of the absolute differences between the observed values and the predicted values. This metric has an important role in assessing the prediction accuracy of a regression model.

The following formula calculates MAE:

$$\mathrm{MAE}(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N}$$

Where, n is the number of observations, y_i is the true values and, $\hat{y}_i$ is the predicted values. The lower the MAE value, the closer the predictions of the model are to the true values and the higher the model's prediction accuracy. R-squared is a metric that measures the fit of the regression model and how much of the variance of the dependent variable it explains. R-squared takes a value between 0 and 1; the closer it is to 1, the better the model describes the variability in the dependent variable.

4. Results and Discussion (Sonuçlar ve Tartışma)

In this section, we evaluated the result of regression and classification analysis of 4 developing countries concerning match achievement.

4.1 Pre-processing Application (Ön işlem Uygulaması)

This data set was finally split into four data sets for four countries, and CNT was removed from these new data sets in a first stage of pre-processing. The number of observations for each of the countries is shown in Table 2 below.

Table 2. *Countries Distribution* (Ülke Dağılımı)

Country	Türkiye	Bulgaria	Mexico	Thailand
Observation	6890	5294	7299	8633

How to deal with the problem of missing data was the second stage of the pre-processing. Missing data is a serious problem when analysing the new datasets at country level. Missing data are listed in Table 3.

Table 3. *Missing Values*

Country	Türkiye	Bulgaria	Mexico	Thailand
ST004D01T	0	0	0	0
ESCS	35	181	824	51
WEALTH	37	163	831	50
HOMEPOS	33	126	824	47
CULTPOSS	121	318	855	59
HEDRES	71	246	844	53
ST011Q02TA	92	225	861	64
MISCED	55	177	845	60
FISCED	57	283	974	82
MMINS	353	1783	3391	469
TMINS	967	2154	4740	4344
PERCOOP	323	1632	3550	353
PERCOMP	286	1560	3129	299
BELONG	103	967	1777	140
EMOSUPS	215	1423	3027	276
PERFEED	124	534	925	113
SUM	2872	11772	27397	6460

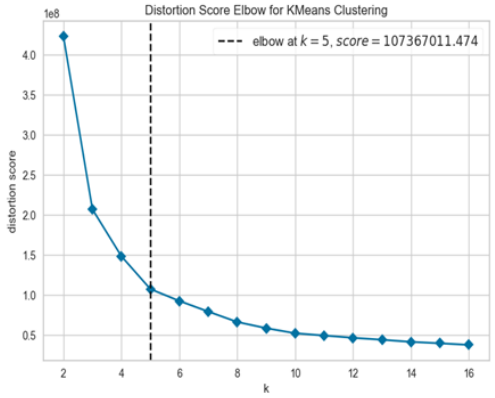
Reviewing Table 3 reveals that simple imputation techniques, like replacing missing values with the arithmetic mean or median, are ineffective. These approaches can render the models useless in both regression and classification, as they tend to emphasize average values and overlook the extremes in the data. Instead, it is crucial to address the data gaps while preserving the range covered by the original dataset. In such cases, either supervised or unsupervised machine learning methods can be applied. However, using supervised learning models for imputation could lead to overfitting issues in subsequent prediction tasks. Given these considerations, we opted for an unsupervised approach, specifically the K-Means clustering algorithm, to handle this process. The K-means algorithm works by grouping the available data points into distinct groups (clusters) based on the similarity of their features. For missing values, the algorithm assigns each instance with missing data to the nearest cluster. The missing values are then imputed using the mean or median of the corresponding feature within that cluster. Rather than simply imputing overall averages or removing entries altogether, this method uses the inherent structure in the data to make informed guesses about the missing entries [30]. This was done by first removing from the country datasets observations with one or more missing observations. The before and after information for the countries is shown in Table 4.

Table 4. *The before-and-after information for the countries* (Ülkeler İçin Öncesi ve Sonrası Bilgileri)

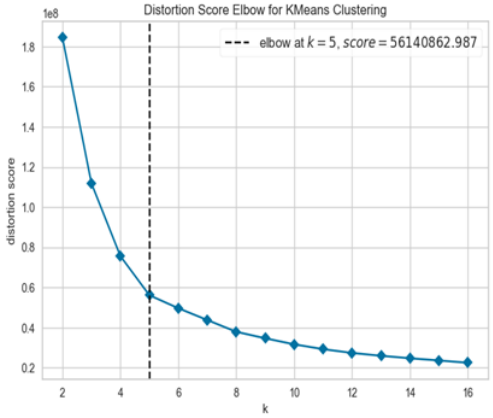
Country	Türkiye	Bulgaria	Mexico	Thailand
n_total	6890	5294	7299	8633
n_missing_values	1466	2848	5096	4555
missing_values_total	2872	11772	27397	6460
n_remains	5424	2446	2203	4078
%	21.3	53.8	69.8	52.8

After presenting the data in Table 4, which outlines the preprocessing of missing data for the countries studied, we applied the K-Means clustering model to datasets of 17 variables with no missing values. This preparatory step is critical for ensuring the integrity and utility of the data before further analysis. As shown in Figure 4.1, the Elbow method was utilized to determine the optimal number of clusters for each dataset, resulting in a division into five clusters for each country. This clustering preserved the original data structure and effectively addressed the gaps caused by missing data.

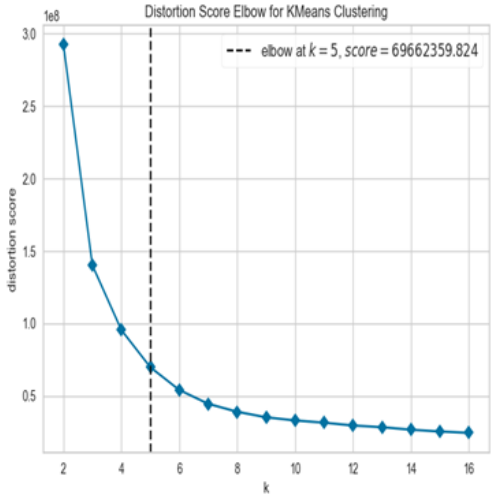
Türkiye



Bulgaria



Mexico



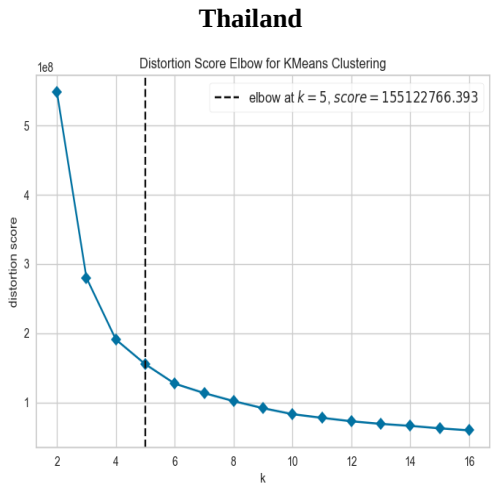


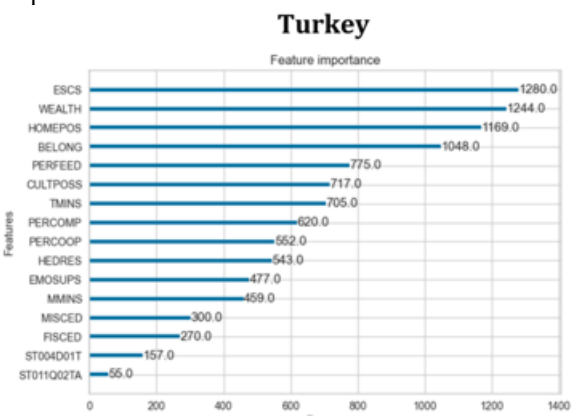
Figure 1. Elbow Method Charts For Countries

(Ülkeler için Dirsek Yöntemi çizelgeleri)

Following the graphical display in Figure 1, it is clear that the Elbow method provides a robust framework for understanding the data structure. By selecting five clusters, we ensure that the data segmentation is neither too sparse to capture essential patterns nor too dense to overfit minor variations. This balance is crucial for the effectiveness of subsequent analytical models which rely on the segmentation quality.

4.2 Feature Selection Results (Özellik Seçimi Sonuçları)

The XGBoost algorithm utilizes F-scores to evaluate the significance of features. The F score is a statistical indicator that measures how much a feature influences the target variable. Features with higher F scores are considered to have a stronger effect on the target variable, whereas lower F scores suggest a weaker influence. In Figure 2, the F scores assigned to each country serve as a key factor in determining the importance of features.



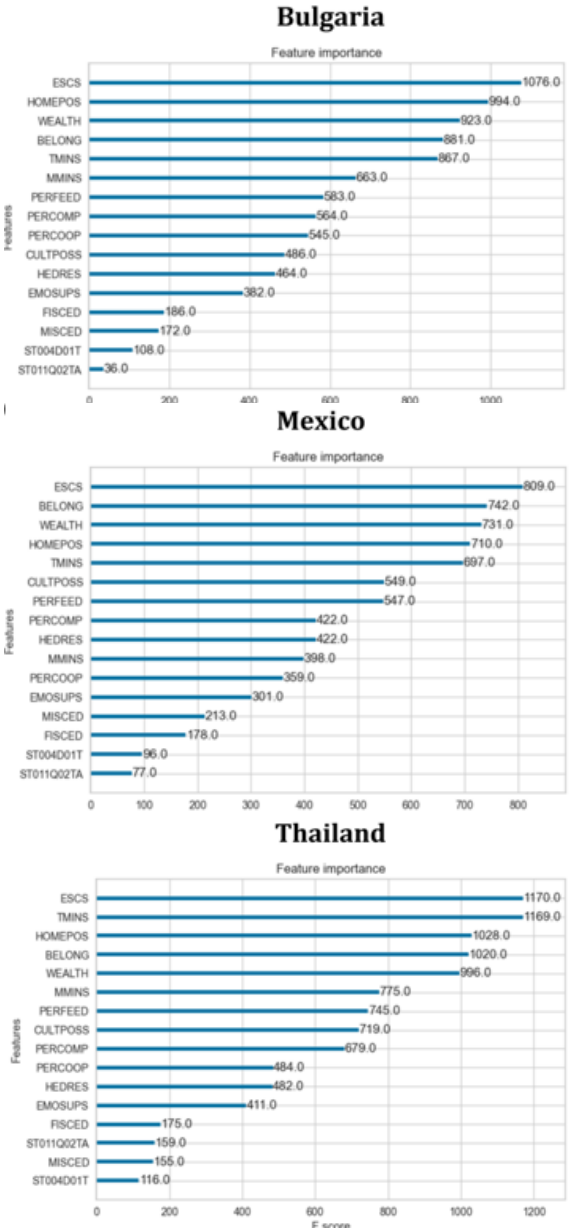


Figure 2. Feature Importance of Variables according to Countries (Ülkelere Göre Değişkenlerin Özellik Önemi)

Each variable has a high F-score and is therefore important in explaining the target variable, according to the results of the XGBoost classification model. (Figure 1).

4.3 Regression Results (Regresyon Sonuçları)

We used Mean Absolute Error (MAE) and R-squared, two important evaluation metrics for regression analysis. Both metrics are standard statistical measures used to quantify and evaluate the performance of regression models.

The MAE and R-squared metrics were used to evaluate the performance of regression models. MAE

measured the closeness of predictions to actual values, while R-squared helped assess the model's fit and the explainability of the dependent variable. Both metrics played an essential role in model selection and comparison and contributed to interpreting the regression analysis results. Accordingly, the results for the relevant metrics for each model are presented in Table 5.

Table.5 Evaluation Metrics for Regression Models (Regresyon Modelleri için Değerlendirme Metrikleri)

Country	Linear Regression		Support Vector R.		Decision Tree R.		Random Forest R.	
	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
Turkey	0.22	57.82	0.18	60.36	0.37	53.01	0.71	50.53
Bulgaria	0.29	59.32	0.32	56.65	0.55	50.00	0.80	47.45
Mexico	0.35	44.34	0.46	36.81	0.55	35.44	0.80	33.29
Thailand	0.43	57.52	0.46	51.37	0.63	44.04	0.84	41.75

The lower the MAE, the closer the model predictions to reality. The explanatory power of the model is also higher the closer the R-squared is to 1. In this context, when the values for Türkiye are analyzed, it is seen that the highest R-squared for Türkiye belongs to Random Forest Regression (71%). Similarly, the lowest MAE value is also observed in Random Forest Regression (MAE=50.54). Accordingly, the most appropriate regression algorithm for Türkiye is Random Forest.

Moreover, when the values for Bulgaria are analyzed, it is seen that the highest R-square value belongs to Random Forest Regression (80%). Similarly, the lowest MAE value is also seen in Random Forest Regression (MAE=47.43). Accordingly, the most appropriate regression algorithm for Bulgaria is Random Forest.

Similarly, the Mexican results show that the highest R-squared value belongs to the Random Forest Regression (80%). As in the other countries, the least MAE value for Mexico is also seen in the Random Forest Regression (MAE=33.29). Accordingly, the most appropriate regression algorithm for Mexico is Random Forest.

Finally, looking at the results of the values for Thailand, it is apparent that the highest R-squared value belongs to the Random Forest Regression (84%) as in the other countries. In addition, the lowest MAE value for Thailand is also seen in Random Forest Regression (MAE=41.45). Therefore, it can be concluded that the most favorable regression algorithm for Thailand is Random Forest.

In terms of the above information, it is seen that the regression model with the lowest MAE and the highest R-squared value for all countries is Random Forest Regression. Therefore, Random Forest is the most appropriate regression model for the above countries.

4.4 Classification Results (Sınıflandırma Sonuçları)

As with regression models, evaluation metrics for classification models are reported in this section. In Table 6 below, the performance of two-class models by countries is presented comparatively. Accordingly, Türkiye's performance varies between 68-76%, Bulgaria's performance between 74-89%, Mexico's performance between 83%, and Thailand's performance between 79-89%. The SVC model showed the lowest performance in Türkiye and Mexico, while the Logistic Regression model showed the lowest performance in Bulgaria and Thailand. Although all models performed 83% in the dataset in Mexico, F1, Precision, and Recall values differ from model to model. The 83% success rate in all models is due to more missing data in Mexico compared to other countries. In the Mexican dataset, the ratio of observations with one or more missing data to all observations is approximately 70%.

Table.6 *Evaluation Metrics for Classification Models (with 3-class)* (Sınıflandırma Modelleri için Değerlendirme Metrikleri (3 sınıflı))

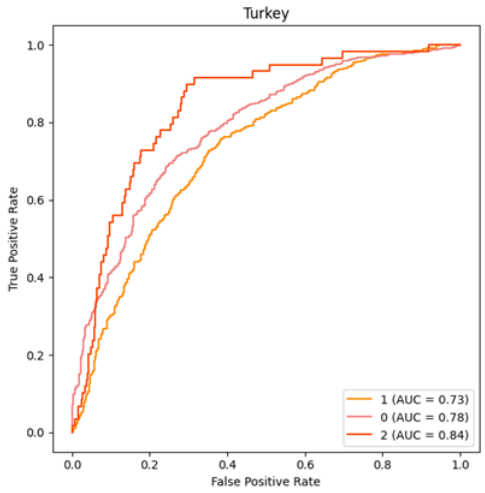
Country	Model	F1	Accuracy	Precision	Recall
Turkey	LR	0.66	0.73	0.71	0.65
	SVC	0.49	0.68	0.72	0.54
	DTC	0.65	0.74	0.74	0.65
	RFC	0.70	0.76	0.75	0.69
Bulgaria	LR	0.65	0.74	0.68	0.64
	SVC	0.72	0.78	0.74	0.71
	DTC	0.76	0.79	0.75	0.77
	RFC	0.86	0.89	0.87	0.85
Mexico	LR	0.56	0.83	0.67	0.55
	SVC	0.45	0.83	0.42	0.50
	DTC	0.60	0.83	0.69	0.58
	RFC	0.64	0.83	0.68	0.62
Thailand	LR	0.74	0.79	0.76	0.73
	SVC	0.80	0.84	0.82	0.80
	DTC	0.84	0.87	0.86	0.83
	RFC	0.86	0.89	0.87	0.85

Following the evaluation metrics presented in Table 6, the ROC curves for Türkiye, Bulgaria, Mexico, and Thailand illustrate the performance of the classification models across different thresholds. Each curve demonstrates the capability of the models to maintain balance between sensitivity and specificity, crucial for predicting the correct class labels. Notably, the ROC curve for Mexico shows a distinctive pattern of a sharp initial rise followed by a stable plateau, indicating a higher initial true positive rate compared

to other countries. These differences underscore the varying performance of the models in each setting, which is further evidenced by the Random Forest model's consistent superiority in handling both high recall and precision levels.

3-Class Model

Türkiye



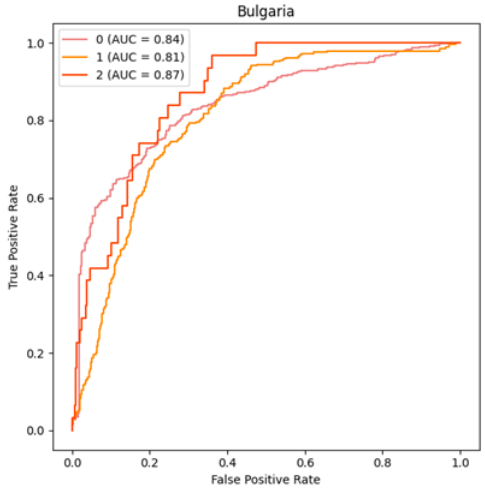
Confusion Matrix

==========

	0.0^A	1.0^A	2.0^A
0.0^P	65.75%	29.90%	4.28%
1.0^P	0.00%	0.07%	0.00%
2.0^P	0.00%	0.00%	0.00%

Note: $class^P$ = Predicted, $class^A$ = Actual

Bulgaria



Confusion Matrix

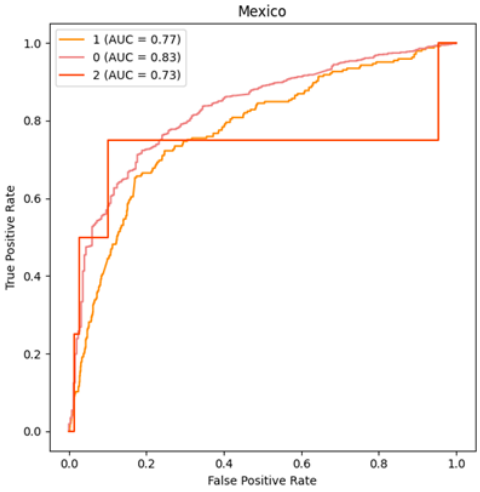
==========

	0.0^A	1.0^A	2.0^A
0.0^P	65.34%	18.79%	1.32%
1.0^P	4.34%	8.59%	1.61%
2.0^P	0.00%	0.00%	0.00%

Note: $class^P$ = Predicted, $class^A$ = Actual

==========

Mexico



Confusion Matrix

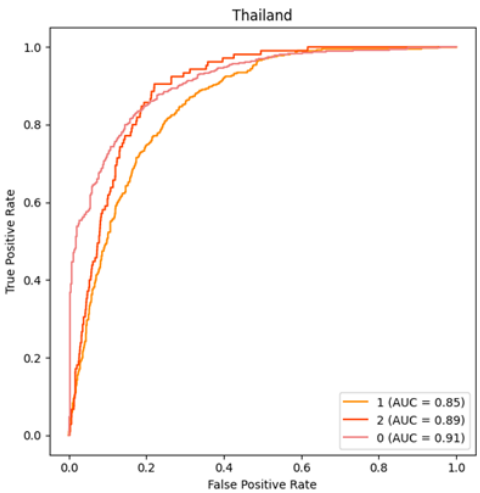
==========

	0.0^A	1.0^A	2.0^A
0.0^P	82.95%	16.78%	0.27%
1.0^P	0.00%	0.00%	0.00%
2.0^P	0.00%	0.00%	0.00%

Note: $class^P$ = Predicted, $class^A$ = Actual

==========

Thailand



Confusion Matrix

	0.0^A	1.0^A	2.0^A
0.0^P	65.61%	9.67%	1.10%
1.0^P	4.52%	14.13%	4.98%
2.0^P	0.00%	0.00%	0.00%

Note: class^P = Predicted, class^A = Actual

Figure 3. ROC Curve For SVC In 3-Class Model
(3 sınıflı modelde SVC için ROC Eğrisi)

After reviewing the ROC curves presented in Figure 3, which illustrate the differing performances of models across various countries, it is crucial to recognize the findings of Bayirli et al. [18]. This research underscores the high accuracy of the Random Forest model in processing data from Thailand among twelve Asian countries. The study identifies significant predictors of mathematical achievement including the economic, social, and cultural status of the student, family welfare, household possessions, sense of belonging, and time allocated for study. These variables highlight the multifaceted nature of educational achievement and point towards areas for targeted educational interventions to enhance outcomes.

5. Conclusion (Sonuç)

In this study, we assessed the efficacy of various machine learning models to predict student accomplishment using PISA 2018 data using regression and classification methods. The classification methods are logistic regression, support vector, decision tree, and random forest, while the regression techniques are multiple linear regression, support vector, decision tree, and random forest. We also assessed the differences in the performance of these models when other countries and educational characteristics were considered. In selecting countries, we considered four countries with similar economic conditions on different continents. These countries are Türkiye, Bulgaria, Mexico, and Thailand, respectively. Our study's principal goals are to assess the performance of machine learning models on PISA 2018 data and investigate potential applications in education. Also, we are planning to study some more different types of PISA achievements in order to propose this model as a kind of decision support system for decision-makers while they are deciding the education policy.

First, the dataset's missing values were located, and the K-means algorithm was used to fill them in appropriately. Each data point is assigned to a cluster by the K-means algorithm, which groups data points into distinct clusters. Data is sorted using this procedure into groups with related qualities. The dataset was split into five separate categories using the K-means technique. These groupings were formed based on similar characteristics and features in the data. At this point, the group to which the rows containing the missing data belonged was established. The missing rows were then filled with the average of the relevant categories after determining which group the missing data rows belonged to.

Also, we utilized the Gradient Boosting algorithm to conduct feature selection among the various variables associated with mathematics achievement based on the findings from the literature research and included in the PISA 2018 dataset. Feature selection is critical in enhancing the model's performance and minimizing the influence of irrelevant variables. Following the feature selection process, it was observed that each variable significantly impacted predicting the mathematics score. Consequently, a set of 16 variables was identified as crucial predictors for accurately forecasting the math scores.

The appropriate algorithms were run once the data became suitable for machine learning models. Our results show that various machine learning models perform well with PISA data for regression and classification analysis. We evaluate our regression model on mean absolute error (MAE) and R-squared metrics. Furthermore, the F1, Accuracy, Precision, and Recall metrics that we used for classification model evaluation were used to assess the classification model success.

In the regression analysis, according to the results of the related models, the Random Forest Regression model achieved the highest R-square values. This result varies between 71% and 84% across countries, while the linear regression model with the lowest explanatory power has R-square values between 22% and 43%. As a result, it can be seen that the model performs better in both explaining the dependent variable and predicting students' performance. The MAE numbers similarly show that Random Forest Regression has the lowest error rate. These findings suggest that Random Forest Regression is the most appropriate regression technique for Türkiye, Bulgaria, Mexico, and Thailand.

The Random Forest Classification model has the highest F1, Accuracy, Precision, and Recall scores in classification analysis. These findings reveal that the Random Forest Classification model outperforms other models in analyzing PISA data and categorizing students. The Accuracy scores of the RF algorithm across countries ranged from 73% to 83%. The Accuracy results of the other algorithms vary across countries. As a result, the Random Forest

Classification model is the most appropriate method for classification analysis in Türkiye, Bulgaria, Mexico, and Thailand.

This research has shown that machine learning models are practical and efficient for studying PISA data. In particular, it was found that the Random Forest Classification and Random Forest Regression models outperformed other models in classifying students and predicting student achievement. These results could provide a more reliable basis for making educational decisions and aid in developing more data-driven and effective educational policies.

Future research can assess machine learning models in greater detail using more extensive and complete data sets. Additionally, a more thorough study of the PISA data can be carried out using various machine-learning algorithms and techniques. Such research can aid in creating more useful educational policies, practices, and initiatives to raise student achievement.

In conclusion, our study has shown how machine learning models can be powerful and helpful in analyzing PISA data. Based on PISA data, the Random Forest Regression and Random Forest Classification models performed the best and offered insightful information to decision-makers- and policymakers in the field of education. This study highlights the significance of making data-driven decisions in education.

References (Kaynaklar)

[1] Niss, M. (1994). Mathematics in society. Didactics of mathematics as a scientific discipline, 13, 367-378.

[2]Popescu, C., and Laura D. (2009). The relationship between the level of education and the Development State of a Country. ŞtiinŃe Economice, 1 (7), 475-480.

[3] Hollands, R. (1990). Development of Mathematical Skills. London: Blackwell Publishers.

[4] Hanif, N., and Noman, A. (2016). Relationship between school education and economic growth: SAARC countries. International Journal of Economics and Financial Issues, 6 (1), 294-300.

[5] Korkmaz, C., and Correia, A.P. (2019). A review of research on machine learning in educational technology. Educational Media International, 56 (3), 250-267.

[6] Sheridan, Kathleen M, David B., Anne P., and Xiaoli W. (2020). Early math professional development: Meeting the challenge through online learning. Early Childhood Education Journal, 48 (2), 223-231.

[7] Kılıçaslan, H., and Yavuz, H.. (2019). PISA sonuçları ile Türkiye'de eğitim harcamaları ilişkisi." Bilgi Sosyal Bilimler Dergisi, 21 (2), 296-319.

[8] Karslı, N., Berberoğlu, G, And Çalişkan, M. (2019). Türkiye'de PISA fen okuryazarlık puanlarını yordayan değişkenler. Uluslararası Bilim ve Eğitim Dergisi, 2 (2): 38-49.

[9]Yüksel, M. (2022). PISA 2018 Araştırma Sonuçlarına Göre Ülkelerin Bileşik PISA Performans Sıralaması.

[10] OECD (2023). Data. Accessed June 23, 2023. https://www.oecd.org/pisa/data/.

[11] MEB. (2019). "PISA-Uluslararası Öğrenci Değerlendirme Programı." Accessed June 18, 2023. http://pisa.meb.gov.tr/www/raporlar/icerik/5.

[12] Mishra, M, Dash, P.B., Nayak, J., Naik, B. & Swain, S.K. (2020). Deep learning and wavelet transform integrated approach for short-term solar PV power prediction. Measurement, 166, https:// doi.org/10.1016/j.measurement.2020.108250.

[13] Jordan, M. & Mitchell, T.M. (2015). Machine Learning: Trends, Perspectives, and Prospects. Science, 349, 255 – 260, https://doi.org/10.1126/ science.aaa8415

[14] Caruana, R. and Niculescu-Mizil, A. (2006). "An Empirical Comparison of Supervised Learning Algorithms". Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, 25-29 June 2006. http:// dx.doi.org/10.1145/1143844.1143865

[15] Lezhnina, O., and Gábor, K. (2022). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. International Journal of Research & Method in Education, 45 (2), 180-199.

[16] Güre, Ö. B., Kayri, M. and Erdoğan, F. (2020). Analysis of Factors Effecting PISA 2015 Mathematics Literacy via Educational Data Mining. Education & Science/Egitim ve Bilim, 45 (202), 393-415.

[17] Saarela, M., Bülent, Y., Mohammed J. Z., and Tommi, K. (2016). "Predicting math performance from raw large-scale educational assessments data: a machine learning approach." JMLR Workshop and Conference Proceedings.

[18] Bayirli, E.G., Atabey, K., and Ersoy, Ö. (2023). An Analysis of PISA 2018 Mathematics Assessment for Asia-Pacific Countries Using Educational Data Mining. Mathematics 11 (6), 1318.

[19] Alzubi, J., Anand, N. and Akshi, K. (2018). Machine learning from theory to algorithms: an overview. Journal of physics: conference series.

[20] Sharma, A., Dinesh B., and Upendra, S. (2017). "Survey of stock market prediction using machine learning approach." 2017 International conference of electronics, communication and aerospace technology (ICECA).

[21] Fama, E.F. (1965). The behavior of stock-market prices. The journal of Business, 38 (1), 34-105.

[22] Ma, J., Theiler, J. & Perkins, S. (2003). Accurate on-line support vector regression. Neural computation, 15 (11), 2683-2703.

[23] Xu, Min, Pakorn Watanachaturaporn, Pramod K Varshney, and Manoj K A. (2005). "Decision tree regression for soft classification of remote sensing data. Remote Sensing of Environment, 97 (3), 322-336.

[24] Li, X.H. (2013). Using "random forest" for classification and regression. Chinese Journal of Applied Entomology, 50 (4), 1190-1197.

[25] Nick, T.G., and Campbell. K.M. (2007). Logistic regression. Topics in biostatistics, 1(1), 273-301.

[26] Chen, P.H., Chih‐Jen L., and Bernhard, S. (2005). A tutorial on ν‐support vector machines. Applied Stochastic Models in Business and Industry, 21 (2), 111-136.

[27] Yadav, S.K., and Saurabh P. (2012). Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification. World of Computer Science and Information Technology Journal (WCSIT), (2), 51-56.

[28] Priyam, A., Gupta R.A., Anju, R., and Saurabh, S. (2013). Comparative analysis of decision tree classification algorithms. International Journal of current engineering and technology, 3 (2), 334-337.

[29] Liu, Y., Yourong, W., and Jian, Z. (2012). New machine learning algorithm: Random Forest. Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3.

[30] Celebi, M. E., Kingravi, H. A., & Vela, P. A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert systems with applications, 40(1), 200-210.