




Determination of Variables Affecting Reading Skills Using the Boruta Algorithm in a Turkish Sample from the PISA 2018

ARTICLE TYPE	Received Date	Accepted Date	Published Date
Research Article	02.21.2023	01.10.2024	07.25.2024

Sanem Şehribanoğlu¹ 
Van Yuzuncu Yil University

Abstract

This study is all about finding out what makes a difference to how students are grouped according to their reading skills. To do this, we looked at the scores from the reading skill test and put students into groups based on whether they have high or low reading skills. In studies with lots of different variables, feature selection helps us to decide which one is the most important. This means we can reduce the size of the data and remove any variables that don't really matter. In this study, we used the Boruta algorithm, which is based on the Random Forests method. This is one of the feature selection methods, and it's really useful because it can identify shadow features, which are created by taking into account the variables. The Boruta algorithm is one of the wrapper algorithms, which can classify large data in a short time and with high accuracy. In this study, we used the PISA 2018 Türkiye data to identify 27 factors that affect reading skills. These include things like school type, career expectations, socio-economic status, interest and familiarity with ICT, metacognition strategies, and more. We collected and evaluated these factors under subject headings.

Key words: PISA 2018, Türkiye, reading skills, attribute selection, Boruta algorithm.

Citation: Şehribanoğlu, S. (2024). Determination of Variables Affecting Reading Skills Using the Boruta Algorithm in a Turkish Sample from the PISA 2018. *Ankara University Journal of Faculty of Educational Sciences*, 57(2), 655-701. <https://doi.org/10.30964/auebfd.1254457>

1.¹*Corresponding Author:* Assist. Prof. Dr., Economics and Administrative Sciences, Department of Economics, Statistics Department E-mail:sanem@yyu.edu.tr, <https://orcid.org/0000-0002-3099-7599>

Since 2000, the Programme for International Student Assessment (PISA) has evaluated the mathematics, science literacy, and reading skills of 15-year-old students attending formal education. PISA data are useful in identifying deficiencies in education systems and determining student contributions to society (Akkoyunlu & Kurbanoglu, 2003). OECD countries greatly benefit from PISA data. PISA's large number of variables and large-scale data is an important source of information for researchers. Additionally, it offers insight into the education systems of participating countries and allows them to track their progress over the years.

Growth and change in technology have increased reading skills and a diversification of social and cultural changes. In the 90s, reading materials were printed on paper, but different from what older generations are accustomed to, today's users read texts online (OECD 2016). The digital world provides access to all types of information that people want to know. The current situation has led to an increase in students accessing digital resources on the internet instead of printed resources in libraries to obtain information (Bana 2020; Mushtaq et al., 2020; Park & Ranasinghe 2021). According to Hootsuite (2021), Google.com, a search engine, is the most visited website. Readers tend to search for answers to their questions on digital search engines rather than in written sources. However, it is up to the individual to determine which of the thousands of answers is scientifically accurate (Suna et al., 2019). Reading in the digital world encompasses a variety of activities such as conducting online research, reading news articles, sending emails and text messages, applying for jobs, completing forms, using social networking applications, performing banking transactions, and shopping. It is important to approach each of these activities with a clear and objective mindset, using precise language and avoiding biased or emotional language. In addition, it is crucial to maintain a formal register, adhere to conventional structure, and ensure grammatical correctness.

Pont and Werquin (2001) state that the ability to read and write has become essential due to global information flow. This has led to improvements in cognitive, information processing, and computer skills. The definition of literacy varies according to the needs of individuals in different time periods. Literacy is defined as the skills required for basic needs, including reading, writing, and arithmetic. Today, 'basic skills' are commonly referred to in terms of literacy, language, and numerical skills, and 'life skills' in terms of literacy, numerical, and information and communication technology (ICT) skills (UNESCO, 2013).

Rapidly developing technologies and a changing world require people who can adapt to this speed and change. Policy makers and educators who help cultivate future individuals need to raise students who can meet these needs. The OECD has designed computer-based assessments to measure reading proficiency, which were first introduced in the PISA 2009 and then used again in the PISA 2015. The emergence of ICT impact factors was facilitated by this innovation (Xiao et al., 2019). Furthermore, the OECD incorporated the global competence area into PISA 2018 to investigate the cultivation of individuals with global competence in response to

current issues (Mostafa, 2021). According to the OECD (2018), students who possess global competence should be capable of analysing global and intercultural issues, communicating successfully and respectfully with diverse individuals and taking necessary actions for sustainability and collective well-being. However, Cobb and Couch (2021) argued that the study is inadequate and requires further development. Engel et al. (2019) expressed concerns regarding the measurement of global competence at the universal and local levels. They argued that the concept of values was ignored in PISA 2018, despite potentially positive results. However, according to the OECD (2018), students who were tested for global competence were also tested with reading texts, making it possible to measure the relationship between reading skills and global competence.

According to Children's Foundation (2006), family, school, and environment are effective social institutions for acquiring reading habits. Numerous researchers have examined the factors that influence reading skills in this multidimensional process. The table in Appendix 1 presents some studies on reading skills using PISA data in the literature. In this context, variables such as gender, socioeconomic status, parental education levels and occupations, parental and teacher support, and ICT use are important for students.

Recent advances in technology have not only transformed our reading habits but also facilitated computer applications which can analyse large-scale data quickly. In this context, PISA datasets, which contain multidimensional data, can be analysed using data mining (DM) methods to reveal hidden information (Kaunang and Rotikan 2018; Salal et al., 2019), resulting in successful outcomes. The use of DM methods to train data is referred to as educational data mining (EDM) (Asif et al., 2017; Firat & Koyuncu, 2020). According to Jalota and Agrawal (2019), EDM employs educational data and student information to improve evaluation. EDM uses statistics, machine learning, and data mining (Mahajan & Saini, 2020) to develop new algorithms or models by identifying new patterns in data (Romero & Ventura, 2020). The table in Appendix 2 presents studies from the literature and their corresponding methods in this context.

Not all variables in large datasets are relevant to the classification investigated in this study. Dealing with large datasets not only slows down the algorithm but also has disadvantages, such as excessive resource consumption and inconvenient methods. Therefore, researchers prefer small datasets that provide the best possible classification results (Kursa and Rudnicki, 2010). Researchers can manually select variables to work with small data. However, manual variable selection can introduce bias and waste time. It is more advantageous to use feature selection algorithms based on the statistical assumptions found in the literature. When comparing the performance and accuracy of feature selection algorithms in machine learning (ML) studies, it is important to consider the process of selecting and finding useful variables (attributes) in the dataset. Zaffar et al. (2017) found that random forest (RF) performed better than other algorithms in their study, which included 6 feature selection and 15

classification algorithms using various student datasets within the scope of EDM. Jalota and Agrawal (2021) used various feature selection algorithms (correlation and wrapper-based) and classifiers (Bayes-Net, JRip, J48, NB, oneR, RF, SMO, Simple Logistic). They determined the correlation feature properties of SMO and J48, and their selection algorithms had the highest accuracy criteria. Although RF did not provide the best performance, it still provided successful results. Gajwani and Chakraborty (2021) employed various algorithms, including DT, LR, NB, boosting, bagging, voting, and RF, to identify the subset of features that have a direct impact on student grades. Their study found that boosting, voting, and RF produced the most accurate results. In his study, Delen (2010) employed various algorithms, including ANN, DT, LR, support vector machine (SVM), RF (bagging), boosted trees, and information fusion (IF), to identify the factors contributing to attrition among at-risk students. The results indicate that the most accurate algorithms were IF and RF (bagging), in order of importance. Han et al. (2019) examined students' problem-solving skills. Son et al. (2020) used the RF algorithm to identify the variables that affect students' reading skills. Anand et al. (2021) employed Boruta for feature selection and NB and DT as classifiers in EDM.

This study focuses on identifying the factors that affect the reading skills of Turkish students. The aim of this study was to identify important variables that affect students with high and low scores in reading skills and to offer new insights to educators and policy makers. In the first stage of the study, the Boruta algorithm based on RF was used to determine the variables affecting reading skills. The Boruta algorithm was used in the second step to determine the prominent questions using questions that comprise these variables. Petko et al. (2017) noted that possible interpretations of the findings should be carefully considered because the PISA survey items are based on general and student statements. In the third stage of the study, probability tables were created for the questions and variables obtained. The study's final section evaluates the findings and discusses the significant variables for student reading skills based on the results obtained from the PISA 2018 Turkish sample.

Method

Research Model

In large-scale evaluations, the score for the sample is considered rather than the score of the individual student (Pejic & Molcer, 2019). These scores are expressed as plausible values (PVs). Arıkan et al. (2020) stated that methods and software that consider the sample weights of large-scale test data (such as PISA, PIAAC, and TIMSS) should be used. Tat et al. (2019) also noted that in large-scale applications, plausible values will not represent the individual scores of students due to the hierarchical structure of the data (student, school, region, country, etc.). According to Özkan (2021), PISA plausible values should not be used as individual values for students. However, PISA evaluates students in a country using the levels determined by it, considering their PVs (classifying them on the basis of gender, schools,

regions). In the PISA Turkey (2000 - 2018) report, Yilmaz et al. (2020) stated that a student placed at a certain point on the PISA skill scale can perform tasks at that level and below that level. According to PISA reports, PVs cannot be used for the evaluation of individual students. Instead, they are assessed on the basis of whether they meet a certain level. As there is no clear success score for students that can be directly used in PISA's performance predictor studies, PVs are used in EDM studies as they provide general information about students (Bezek Güre et al., 2020; Depren and Depren, 2021; Dong and Hu; Gamazo and Martínez-Abad, 2020; Koyuncu and Gelbal, 2020; Özkan, 2021; Pejic and Stanic Molcer, 2018; Son et al. 2020). In this study, the PVs of students were deemed appropriate for evaluation, as the analysis is based on classification rather than individual scores.

Data Collection Tools

This is a quantitative relational study that uses the student survey of the PISA 2018 Turkey sample (OECD 2019c) as a data collection tool. The scores obtained from the students' reading skill test (PV1-PV10) were averaged, as used in some studies in the literature (Bezek Güre et al., 2020; Dong & Hu, 2019; Lee, 2018; Pejic & Stanic Molcer, 2018) and in PISA final reports. No special permission was required for the use of the data because it is an open-access dataset.

Ethical Committee Approval

The study utilized the data set made available to all researchers by OECD (2019b). Therefore, there is no need for an ethical declaration for the study.

Data Analysis

Feature Selection (FS)

FS is a widely used technique in pattern recognition. It involves removing noisy or unnecessary variables from the data to make it more comprehensible (Venkata and Lingamgunta, 2020; Yan and Zhang, 2015). This reduces data storage requirements, training and application times (Guyon & Elisseeff, 2003), and improves algorithm speed and prediction accuracy (Guyon & Elisseeff, 2006).

Random Forest

The RF algorithm is a non-parametric statistical method developed by Breiman (2001). It can be used for classification, feature selection, and regression (Lahouar and Slama, 2015). RF is an ensemble learning method that combines the ideas of clustering and bootstrap by creating a larger tree with better performance by merging a series of decision trees (Unpingco, 2019). This is a powerful method used in regression and classification problems (Bezek Güre et al., 2020).

Boruta Algorithm

This wrapper method utilizes the RF classification algorithm (Ahmed et al., 2021; Kursu & Rudnicki, 2010) to determine feature importance. The algorithm expands the dataset by multiplying it and uses the Z-score obtained from the random

forest (Iman & Ahmad, 2020). The highest Z-score among the shadow features is then used as a reference to determine which features are important or unimportant (Kursa & Rudnicki, 2010). Important attributes are those above the reference, whereas those below are considered unimportant. The analysis is then repeated with the removal of unimportant variables until only the important features are retained.

In studies related to EDM, classifications are made by considering different features. Asif et al. (2017) classified student performances into successful and unsuccessful groups, while Kasap et al. (2021) coded the first three levels of PISA reading skill scores, which are divided into six levels, as low and the other three levels as high. Emdadi and Eslahchi (2021) and Xu et al. (2019) classified based on the median value of the dependent variable in their studies. They calculated the average reading skills scores (PV1-PV10) of 6890 students (MPVREAD). To improve classification performance and make the study easier to understand, the dependent variable MPVREAD was classified as in previous studies by Asif et al. (2017), Mahajan and Saini (2020), Son et al. (2020), and Han et al. (2020), who employed a sensitive average of the arithmetic mean. The median value of MPVREAD (MedPVREAD = 468) was calculated. Results below the median value were recoded as 1 (low), and averages equal to or above the median value were recoded as 2 (high) (PVREAD).

The variables used in the PISA tests are categorized into three groups: simple indices, scale indices, and trend scale indices. Simple indices are obtained through arithmetic transformation and recoding, whereas scale indices are created from parameters obtained by scaling more than one item (OECD 2019b). Table 1 provides a list of indices used in the study.

Indices with missing or undefined information were removed from the dataset. The deleted indices include ATTIMMP, BODYIMA, CHANGE, CNTSCHID, CNTSTUID, CURSUPP, EMOSUPP, EFFORT1, EFFORT2, EMOSUPS, FCFMLRTY, FLCONFIN, FLCONICT, FLFAMILY, FLSCHOOL, GCAWAREP, INFOCAR, INFOJOB1, INFOJOB2, INTCULTP, JOYREADP, LANGFATHER, LANGFRIEND, LANGMOTHER, LANGSCHMATES, LANGSIBLINGS, LANGTEST_COG, LANGTEST_PAQ, LANGTEST_QQQ, OCOD1, OCOD2, OCOD3, OECD, PASCHPOL, PQSCHOOL, PRESUPP, PROGN, SCCHANGE, SOCONPA, STUBMI, and SUBNATIO. The analysis included 86 variables, one of which (PVREAD) was dependent. Gender was also considered. The Random Forest (Liaw and Wiener, 2018) and Boruta (Kursa 2020) packages in the R program were used for the interpolation method and analysis of missing data. Table 1 lists the 84 index names used in the first phase of this study. In both stages, the evaluation was based on the Z transformation (meanImp) values. In the first stage of the application, 27 significant variables (indices) were obtained from 100 trees, including indices that affect reading skills (PVREAD), with an accuracy value of 69.7.

Table 1.
Indices Used In The Study

Student Indices
ADAPTIVITY, AGE, ATTIMM, ATTLNACT, AWACOM, BEINGBULLIED, BELONG, BFMJ2,BMMJ1, BSMJ, COBN_F, COBN_M, COBN_S, COGFLEX, COMPETE, CULTPOSS, DIRINS, DISCLIMA, DISCRIM, DURECEC, EMOSUPS, ESCS, EUDMO, FISCED, FISCED_D, GCAWARE, GCSELFEEFF, GFOFAIL, GLOBMIND, GRADE, HEDRES, HISCED, HISCED_D, HISEI, HOMEPOS, , ICTRES, IMMIG, INTCULT, ISCED, ISCEDD, ISCEDL, ISCEDO, JOYREAD, LANGN, LMINS, MASTGOAL, METASPAM, METASUM, MISCED, MISCED_D, MMINS, PARED, PAREDINT, PERCOMP, PERCOOP, PERFEED, PERSPECT, PISADIFF, REPEAT, RESILIENCE, RESPECT, SCREADCOMP, SCREADDIFF, ,SMINS, ,STIMREAD, STRATUM, SWBP, TEACHINT, TEACHSUP, TMIN, UNDREM, WEALTH, WORKMAST
ICT Indices
ICTHOME, ICTSCH, ENTUSE, USESCH, AUTICT, INTICT, COMPACT, SOIAICT, ICTOUTSIDE, ICTCLASS, HOMESCH

During the second phase of the study, 27 indices were analysed by asking related questions. As questions could not be obtained regarding the BSMJ (expected occupational status) index, it was added to the analysis. The results of the analysis showed that out of the 159 variables determined, 52 variables (problems) were highlighted with 100 trees and an accuracy value of 67.2.

Results

In the initial stage of the application, Figure 1 displays the first analysis results of 85 variables believed to impact reading skills. The blue box plots indicate the minimum, average, and maximum Z scores of the shadow attributes. The red, yellow, and green box plots represent the Z scores of rejected, uncertain, and approved attributes, respectively. The analysis was repeated by excluding unimportant and uncertain variables. A total of 27 variables (indices) were selected for the study and are listed in Table 2 in order of importance.

Table 2.
Indices That Stand Out In Reading Skills

Indices	meanImp	Indices	meanImp	Indices	meanImp
1.METASPAM	47,77	10.HOMEPOS	26,03	19.SCREADIFF	17,24
2.PISADIFF	33,15	11.SMINS	24,30	20.GCAWARE	17,23
3.METASUM	30,72	12.GCSELFEEFF	24,26	21.BSMJ	17,16
4.ESCS	30,56	13.ICTHOME	20,62	22.COMPICT	16,94
5.USESCH	27,46	14.STRATUM	19,38	23.JOYREAD	14,28
6.ISCEDO	27,33	15.UNDREM	19,29	24.FISCED	13,44
7.INTICT	26,66	16.WEALTH	18,35	25.TMINS	13,33
8.ICTRES	26,53	17.ENTUSE	17,97	26.GLOBMIND	10,35
9.AWACOM	26,04	18.HOMESCH	17,65	27.MISCED	8,43

Upon examining Table 2, it is evident that the indices pertain to the access and use of ICT resources (USESCH, ENTUSE, HOMESCH, ICTHOME), interest in ICT and perceived competence (INTICT, COMPICT), expected professional status (BSMJ), economic social cultural level (ESCS, HOMEPOS, ICTRES, WEALTH, FISCED, MISCED), time devoted to education (SMINS, TMINS), metacognition (METASPAM, METASUM, UNDREM), attitudes toward reading (JOYRED, PISADIFF, SCREADIFF), curiosity about the world and other cultures (GLOBMIND, AWACOM, GCSELFEEFF, GCAWARE), and the type of school the student attends (STRATUM, ISCEDO). These indices are grouped under nine headings.

During the second phase of the study, 52 questions were identified as relevant after analysing 159 questions grouped under 9 headings (refer to Figure 2). These questions are listed in Table 3, ordered by index name and importance. Statistically significant relationships were found between the reading skill score averages (MPVREAD) and the prominent indices, except for STRATUM, as shown in Table 4. The most significant relationships were observed for METASPAM, METASUM, and PISADIFF, which were among the prominent indices in the first stage (Table 2).

Figure 1.
First-Stage Analysis Results

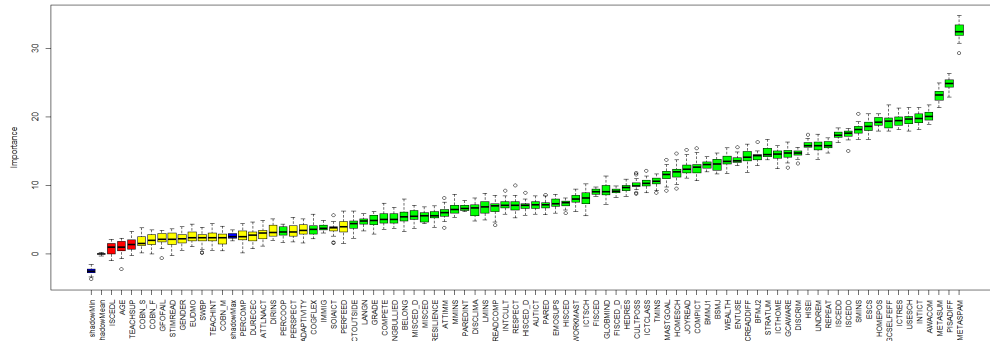


Table 3.
Prominent Questions And Indices In Reading Skills

Variables	Index	meanImp	Variables	Index	meanImp
ST013Q01	HOMEPOS	31.19	ST011Q07	HOMEPOS	13.60
ST166Q03	METASPAM	29.63	ST007Q01	FISCED	13.48
ST166Q02	METASPAM	24.18	IC008Q04	ENTUSE	13.36
ST059Q03	SMINS	23.60	IC010Q04	HOMESCH	13.29
IC008Q08	ENTUSE	22.74	IC001Q06	ICTHOME	13.13
ST163Q02	PISADIFF	22.40	ST164Q05	UNDREM	11.94
ST002Q01	STRATUM	22.24	IC008Q07	ENTUSE	11.80
IC011Q05	USESCH	22.13	IC008Q09	ENTUSE	11.68
BSMJ	BSMJ	18.29	ST166Q04	METASPAM	11.65
ST060Q01	TMINS	18.14	ST165Q02	METASUM	11.60
ST166Q05	METASPAM	18.07	ST160Q04	JOYREAD	11.38
IC001Q10	ICTHOME	18.04	IC010Q07	HOMESCH	11.00
ST196Q04	GCSELF EFF	18.04	IC013Q01	INTICT	10.90
IC008Q05	ENTUSE	17.54	IC013Q13	INTICT	10.87
ST012Q05	HOMEPOS	17.39	ST163Q04	PISADIFF	10.33
IC014Q06	COMPICT	17.27	ST005Q01	MISCED	9.93
IC013Q04	INTICT	17.13	ST161Q08	SCREADDIFF	8.24
ST197Q01	GCAWARE	17.03	ST161Q07	SCREADDIFF	7.94
IC011Q09	USESCH	16.80	ST166Q01	METASPAM	7.24
IC011Q06	USESCH	16.79	ST164Q04	UNDREM	6.78
ST165Q04	METASUM	16.03	ST164Q01	UNDREM	6.61
IC011Q08	USESCH	15.59	ST165Q03	METASUM	5.43
IC008Q12	ENTUSE	15.28	ST164Q03	UNDREM	5.28
ST165Q05	METASUM	14.90	ST164Q02	UNDREM	4.67
ST163Q03	PISADIFF	14.06	ST165Q01	METASUM	4.48
ST196Q02	GCSELF EFF	14.03	ST164Q06	UNDREM	2.99

Figure 2.
Second-Stage Analysis Results

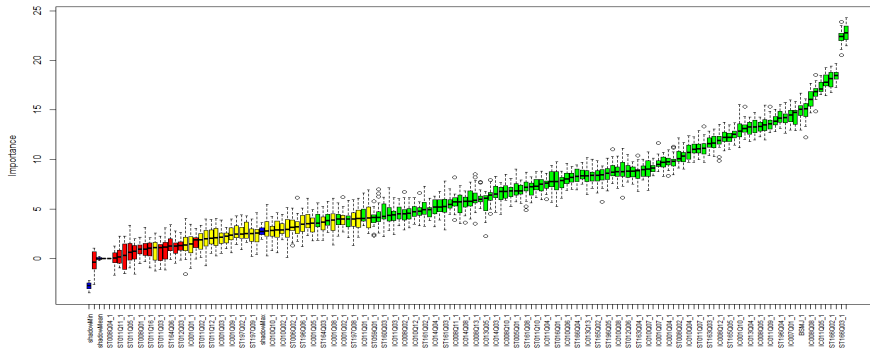


Table 4.
MPVREAD And The Index Correlation Table

	METASPAM	METASUM	PISADIFF	HOMEPOS	UNDR EM
MPVREAD	.423**	.371**	-.356**	.337**	.294**
	GCESELFEFF	BSMJ	SCREADDIFF	JOYREAD	SMIN S
MPVREAD	.269**	.259**	-.252**	.237**	.231**
	COMPICT	FISCED	USESCH	INTICT	GCA WARE
MPVREAD	.206**	.198**	-.192**	.191**	.187**
	MISCED	ENTUSE	ICTHOME	HOMESCH	TMIN S
MPVREAD	.187**	.124**	.115**	-.064**	.041**

** . Correlation is significant at the 0.01 level (2-tailed).

Table 5.
Correlation Table Of Reading Skills

	MPVMATH	MPVSCI
MPVREAD	.872**	.922**
MPVMAT	1	.918**
H		

Attitudes Toward Reading

It was discovered that the indices for enjoyment of reading (JOYRED), perception of difficulty of the PISA test (PISADIFF), and perception of difficulty performing reading tasks (SCREADDIFF), which were included in the heading of attitudes toward reading determined in the first stage, had an effect on students' reading skills. Table 2 shows that PISADIFF ranked sixth in terms of importance among the three indices. Table 4 shows the statistically significant negative relationships between SCREADDIFF and PISADIFF with PVREAD.

Metacognition

Statistically significant relationships were found between the understanding and remembering (UNDREM), summarizing (METASUM) and assessing credibility (METASPAM) indices and PVREAD. These relationships were used to evaluate the students' cognitive states regarding reading, as shown in Table 4.

According to the classification of reading ability (refer to Table 2), METASPAM and METASUM are among the prominent indices in the priority ranking. No changes in content were made to ensure that the improved text was as close as possible to the source text. After analysing the results, it was confirmed that reading strategies, namely METASPAM, METASUM, and UNDREM, play a crucial role in classifying reading skills.

Time Devoted to Training

Although there was a statistically significant relationship between the weekly lesson hours for the students in science (SMINS) and total courses (TMINS) on reading skills (Table 4), the correlation value of TMINS was almost zero ($r = 0.041$).

To observe the relationship between reading skills and science knowledge, we examined the average scores for reading (MPVREAD), mathematics (MPVMATH), and science (MPVSCI) skills (Table 5). The analysis revealed a strong correlation ($r = 0.92$) between the average reading skill (MPVREAD) and the average science skill.

Knowledge about the World and Other Cultures

During the first phase of the study, four indices (GLOBMIND, AWACOM, GCSELFEFF, GCAWARE) were identified as being related to curiosity about the world and other cultures (see Table 2). However, in the second phase of the study, only two of these indices were found to be relevant. GCSELFEFF measures a student's self-efficacy in explaining or discussing global issues, while GCAWARE assesses a student's awareness of global issues. Table 3 shows that these two indices had a positive and significant correlation with students' reading skills.

Economic-Social-Cultural Level

In the first phase of the study, four indices were identified under the heading of economic–social– cultural level (ESCS, HOMEPOS, ICTRES, WEALTH) (Table 2). ESCS is a composite score consisting of three indices.

ESCS comprises parental occupational status (HISEI), parental education (PAREDINT), and household possessions (HOMEPOS) indices. The HOMEPOS index is calculated on the basis of the indices of wealth (WEALTH), cultural assets (CULPOSS), and home education resources (HEDRES), which include the number of books available at home. The ICTRES index is based on the ICT resources owned. In the second phase of the study, questions representing these three indices - HOMEPOS, FISCED, and MISCED-were highlighted. The FISCED and MISCED indices provide information about the parents in the PAREDINT index. Statistically significant relationships were found between these three indices and reading skills, as shown in Table 4.

Access and Use of ICT Resources

Table 2 presents USESCH, which indicates the frequency of digital device use at school; ENTUSE, which represents usage during out-of-school activities; and HOMESCH, which indicates usage during homework outside of school, all falling under the heading of access and use of ICT resources. ICTHOME are indices obtained for ICT devices used at home.

In Table 3, questions representing these four indices were highlighted in the second phase of the study, and it was confirmed that they were critical to the classification of reading skills. Statistically significant relationships were found between the four indices and reading skills (refer to Table 4). However, only the USESCH and HOMESCH indices had statistically significant negative correlations with PVREAD. It is worth noting that the correlation value for HOMESCH ($r = -0.064$) was very low.

Interest in and Perceived Competence in ICT

The two indices, INTICT and COMPICT, collected under the title of interest in ICT and perceived competence in Table 2, maintained their significance in the second phase of the study. These indices evaluate students' interest in digital media and devices (INTICT) and their perceived competence (COMPICT). Both indices had positive and statistically significant relationships with reading skill averages (Table 4).

Expected Occupational Status

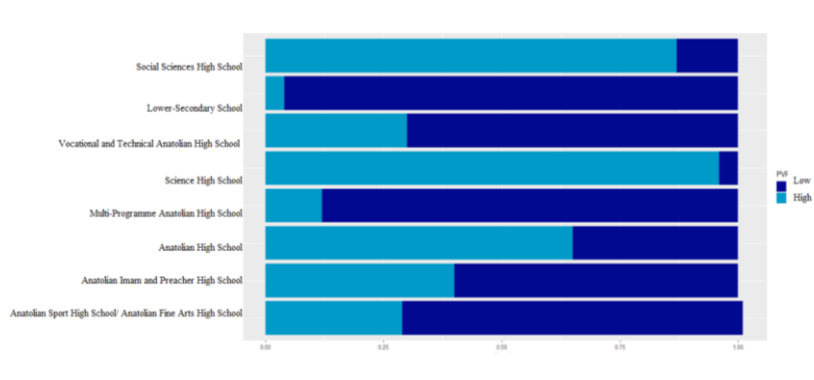
The BSMJ index used in both stages of the study is an edited version of student answers to questions about their desired job at thirty years of age, classified according to the International Labor Organization's (ILO) structure. High scores indicate a

preference for high professional status (OECD, 2020). Table 4 shows the statistically significant positive relationship between BSMJ and reading skills.

School Type

In both stages of the study (Table 2, Table 3), STRATUM was found to be significant. After rearranging the data based on the student school type, they were classified according to their reading skill scores. Figure 3, which takes probabilities into account, was obtained as follows: students in Anatolian, science, and social sciences high schools were found to have high reading skills.

Figure 3.
Reading Skill Levels Based On The School Type



Discussion, Conclusion and Suggestions

The phrase 'knowledge itself is power', coined by Francis Bacon in 1597, remains relevant in today's rapidly advancing technological age. Reading is a crucial means of acquiring knowledge, and in the 21st-century, reading skills are more important than ever. As defined by Fabunmi and Folorunso (2010), the modern and complex world we live in demands strong reading skills for personal and professional success. Nzeyimana and Bazimaziki (2020) suggest that new strategies are required to address poor reading skills, which can lead to reliance on unreliable information and incomplete knowledge due to the rapid technological advancements in today's world.

An individual's effective use of metacognitive strategies significantly impacts the quality of learning (Arslan, 2020). Students who possess metacognitive awareness know which strategies to use, where, and when (Firat and Koyuncu, 2020). Depren and Depren (2021) found that metacognition indices are effective in improving reading skills. This finding is supported by the studies of Firat and Koyuncu (2021) and Son et al. (2020), as well as our own study.

The rapid rise of the digital world has made careful reading essential. According to UNESCO (2013), literacy now encompasses life skills. In many parts of the world, digital reading skills are considered crucial for individuals to achieve their goals and participate in society (OECD, 2016). To enable students to deal with the challenges posed by the rapid advancement of technology, education systems must incorporate these changes directly into classroom education (Henry 2006; Keskin 2014). In addition, students should receive support in the form of digital reading skill training.

Although technology and the internet provide unlimited access to information, a lack of knowledge about research topics may indicate insufficient digital media literacy (OECD, 2018). Students who possess advanced global competence can develop a global perspective in various disciplines such as science, literature, mathematics, biology, and history. This can help them learn in-depth and achieve greater success in their educational pursuits (Tedmem, 2019). Furthermore, the young generation with low global competence may be susceptible to biased and fake news, as noted by the OECD in 2018. Therefore, it is crucial for school administrators and teachers to guide students in using digital resources in a more meaningful and productive manner, while also providing support through metacognitive strategies. This will bring us one step closer to achieving the intended goals of sustainability and collective welfare. Our study aligns with the opinion of the OECD (2018). It was found that students with a cultural background and an interest in global issues tended to have higher reading skills.

Oriogu et al. (2017) stated that social media (e.g. Facebook, Whatsapp, Twitter) is the most significant factor hindering students' reading habits. The UNESCO report (West and Chew, 2014) found that two-thirds of individuals in five developing countries read on their phones (OECD, 2016; West and Chew, 2014). The report also noted that most of these mobile readers were young and more educated than average. Currently, changes in reading habits have increased e-reading among young people. According to Mushtaq et al. (2020), electronic content offers several advantages.

In a study of secondary school students, Adeyokun et al. (2020) found that reading online via smartphones and laptops had a positive impact on reading habits. Srijamdee and Pholphirul (2020) and Vázquez-Cano et al. (2020) reported that the use of ICT during education has a positive impact on academic achievement. However, the use of ICT for non-educational activities has no effect on academic performance. Xiao et al. (2019) found a negative correlation between academic success and the use of ICT for social interaction, whereas Petko et al. (2017) reported similar findings for digital entertainment at home. Gubbels et al. (2020) and Xiao et al. (2019) found that excessive use of ICT can reverse the positive effect on student achievement, as the relationship between ICT use and student achievement follows an inverted U-shape. The authors emphasized that students spend their time is more important than the amount of time they spend on the internet. According to Henry (2006), teachers must comprehend the new literacies that are emerging in their classrooms to prepare students for life. It is crucial not to isolate students from the internet while searching

for solutions. In addition, students should be educated about the critical and qualified use of technology and taught strategies to avoid inappropriate content for their age (Vázquez-Cano et al., 2020). Lee and Wu (2012) found positive correlations between students' perceptions of ICT and their academic performance, as well as between ICT attitudes and reading performance (Ertem, 2021; Petko et al., 2017). Similarly, Hu et al. (2018) reported interest and proficiency in ICT, and Srijamdee and Pholphirul (2020) stated that students who are experienced and familiar with ICT perform well in reading, science, and mathematics. These findings are consistent with those of the existing literature. Students who report an interest in and proficiency with ICT tend to have higher reading skill scores. In today's rapidly advancing technological era, it is important to enhance students' acceptance of ICTs and help them feel competent in this regard, especially considering the smart devices that are now commonly used in homes.

Determining the compatibility of students' career expectations, future employment opportunities, and educational plans is crucial (Sikora and Pokropek, 2006). According to Consulting and Trust (2013), the youth unemployment rate is high among the 18-24 age group who are not in full-time education or employment, and this is linked to low literacy. The aim of this study is to improve the literacy of young people by increasing their future prospects. The 11-14 age group was identified as a special focus point and should be supported, particularly in the first years of secondary education, to create realistic employment demands. As noted by Gamazo and Martínez-Abad (2020), expected professional status is a factor that enhances student performances. It was determined in our study that students with high professional expectations achieved high scores in reading skills.

One of the important factors affecting reading skills is school type (Çelik and Yurdakul, 2020; Ertem, 2021; Fırat and Koyuncu, 2020; OECD, 2019a). Educational practices and curricula can contribute to students' academic success (Flores-Mendoza et al., 2021). According to Suna et al. (2020), the difference in success performance between schools has been a long-discussed issue in Türkiye, starting from secondary education and continuing to increase. Science high schools have the highest average score in terms of reading skills, as reported in previous PISA results (Yılmaz et al., 2020). To reduce disparities between different types of schools, policymakers and educators should consider providing support to low-achieving schools (Suna et al., 2020) and reviewing their curricula.

The academic performance of students is influenced by various factors, including the socioeconomic background of their parents (Fırat and Koyuncu, 2020; Geske and Ozola, 2008; Hu et al., 2021). Son et al. (2020) discovered that Korean students with low socioeconomic indicators had lower reading skill scores in PISA 2018. Chung et al. (2021) discovered a positive correlation between socioeconomic indicators and reading skills among Finnish and Korean students. Similarly, Suna et al. (2019) reported a positive but low impact on their Turkish sample. These findings align with our study results.

Family plays a crucial role in a child's life. Parental education has a significant impact on occupation and income, as well as the child's reading ability (Geske and Ozola 2008; Le et al. 2019; Rojas-Torres et al. 2021; Son et al. 2020; Vázquez-Cano, De la Calle-Cabrera, et al. 2020). Our findings highlight the importance of parental education as a variable.

In our study, science and total weekly course hours were found to be the primary factors affecting reading skills, while Dong and Hu (2019) identified the time spent on students' test language (LMINS) as one of the indices. Similarly, Nadaf et al. (2021) found that total learning time (TMINS) was the most significant feature contributing to students' cognitive success after excluding other course periods from the analysis. Chung et al. (2021) noted that training for more than 35-hour per week had a less positive, or even negative, impact on the results. The study found that while TMINS was an important variable for Korean students, activity sub-headings such as free time and independent reading time were more prominent in the learning times of Finnish students. Depren and Depren (2021) found that the amount of time Turkish students spend on science subjects had a statistically significant impact on their reading skills. Our study also found a strong positive correlation between reading skills and science literacy.

From face-to-face classrooms where traditional education systems are hosted, with today's technological developments, the relationship between ICT and education has become very important, albeit due to necessity. Training started with online courses and videos, and especially with COVID-19 in 2020, education had to be provided in virtual classes. To enhance students' reading skills in this era of changing reading culture, it is crucial to concentrate on metacognitive strategies and ICT literacy. This will enable students to safeguard themselves against negative online experiences. Policy makers and educators should provide students with increased access to the internet and digital devices and teach them how to use them effectively. This will help prepare the younger generation to keep pace with rapidly evolving technology. Efforts should be made to reduce disparities between schools. Rather than opposing evolving reading habits, we should adapt to them and utilise technological resources that can enhance students' learning, such as audiobooks, podcasts and educational videos, to encourage more engagement. Furthermore, conducting additional studies that consider genders and school types, not included in this research, may yield valuable insights for policymakers and educators. These studies should examine the variables that are prominent in reading skills.



PISA 2018 Türkiye Örneğinde Okuma Becerisini Etkileyen Değişkenlerin Boruta Algoritması ile Belirlenmesi

MAKALE TÜRÜ	Başvuru Tarihi	Kabul Tarihi	Yayın Tarihi
Araştırma Makalesi	21.02.2023	10.01.2024	25.07.2024

Sanem Şehribanoğlu¹ 
Van Yüzüncü Yıl Üniversitesi

Öz

Bu çalışmanın amacı öğrencilerin okuma beceri düzeylerine göre sınıflandırma üzerinde etkisi olan değişkenlerin araştırılmasıdır. Bu amaçla öğrencilere ait okuma beceri testinden aldıkları puanlar (PV) dikkate alınmış, yüksek ve düşük okuma becerisine sahip olarak belirlenen sınıflandırma grupları oluşturulmuştur. Çok sayıda değişkene sahip olan çalışmalarda hangi değişkenin daha etkin olduğunu tespit etmek için kullanılan öznitelik seçimi verilerin boyutlarının azaltılmasını ve ilgisiz değişkenlerin çıkarılmasını sağlar. Bu çalışmada da öznitelik seçim yöntemlerinden biri olan Rastgele Ormanlar yöntemini temel alan Boruta algoritması kullanılmıştır. Z-skorları hesaplanan bu algorithmada değişkenler dikkate alınarak oluşturulan gölge öznitelikler referans olarak dikkate alınır. Sarmalayıcı algoritmalarından biri olan Boruta algoritması, büyük verilerde kısa sürede ve yüksek doğrulukla sınıflandırma becerisine sahiptir. PISA 2018 Türkiye verilerinin kullanıldığı bu çalışmada sınıflandırılmış okuma beceri değerini etkileyen 27 indis belirlenmiştir. Belirlenen bu değişkenler Okul türü, kariyer beklentisi, sosyo-ekonomik durum, BİT'e olan ilgi ve aşinalık, üst biliş stratejileri vb. konu başlıkları altında toplanmış ve değerlendirilmiştir.

Anahtar sözcükler: PISA 2018, Türkiye, okuma becerisi, öznitelik seçimi, Boruta Algoritması.

¹*Sorumlu Yazar:* Dr. Öğr. Üyesi, İktisadi ve İdari Bilimler , İktisat Bölümü , İstatistik Bölümü E-mail:sanem@yyu.edu.tr , <https://orcid.org/0000-0002-3099-7599>

Uluslararası Öğrenci Değerlendirme Programı (PISA), 2000 yılından bu yana 15 yaş grubu örgün eğitime devam eden öğrencilerin matematik, fen okuryazarlığı ve okuma becerilerini değerlendiren uluslararası bir çalışmadır. PISA öğrencilerin sosyal yaşama ve topluma katkılarının belirlenmesinde faydalıdır (Akkoyunlu ve Kurbanoglu, 2003). OECD ülkeleri eğitim sistemlerindeki eksikleri fark edebilmek amacıyla PISA verilerinden faydalanır. PISA'nın çok sayıda değişken ve büyük ölçekli verilere sahip olması araştırmacılar için önemli bir bilgi kaynağıdır. Aynı zamanda katılımcı ülkelerin eğitim sistemleri hakkında bilgiler sunmasının yanı sıra yıllar içinde ülkelerin kendi değişimlerini takip edebilmelerine de olanak sağlaması açısından önemlidir.

Okuma becerisi, teknolojinin hızla büyümesi ve değişimi ile hem toplum hem de kültürel değişimleri artırmakta ve çeşitlendirmektedir. 90'lı yıllarda okuma materyalleri kâğıtlar üzerine basılırken günümüz kullanıcılarının çevrimiçi okuduğu metinler, alışıktığımızdan farklıdır (OECD 2016). Dijital dünya, insanların bilmek istediği her türlü bilgiye erişimi kolaylaştırmaktadır. Bu durum öğrencilerin bilgiye erişim için basılı kaynakların olduğu kütüphaneler yerine internet üzerinden dijital kaynaklara erişim eğilimlerini artırmıştır (Bana 2020; Mushtaq vd. 2020; Park ve Ranasinghe, 2021). Bir arama motoru olan Google.com, Hootsuite'ın (2021) hazırladığı rapora göre web site sıralamasında en çok ziyaret edilen adres olarak belirlenmiştir. Okuyucular bir sorunun cevabını yazılı kaynaklardan çok dijital arama motorlarında aramaktadır. Arama sonucu kullanıcıların karşısına çıkan binlerce cevap arasından hangisinin doğru ve bilimsel olduğunu anlaması bireyin kendi başına vereceği bir karardır (Suna vd., 2019). İnternet sitesinde araştırma yapmak, gazete, e-posta, kısa mesajları okumak, iş başvurusunda bulunmak, form doldurmak, sosyal ağ uygulamalarını kullanmak, bankacılık işlemlerini gerçekleştirmek, alışveriş yapmak dijital dünyanın yeni okuma türleri olarak her geçen gün çeşitlenerek karşımıza çıkmaktadır.

Pont ve Werquin'e (2001) göre küresel bilgi akışı okuma ve yazmayı zorunlu becerilere dönüştürmüştür. Bu durum kullanıcıların bilişsel, bilgiyi işleme ve bilgisayar kullanma becerilerini de geliştirmiştir. Okuryazarlık tanımı, insanların yaşadıkları her dönem ihtiyaçlarına göre değişim göstermektedir. Temel ihtiyaçlar için gereksinim duyulan okuma, yazma ve aritmetik beceriler olarak tanımlanan okuryazarlık, bugün okuma-yazma, dil ve sayısal beceriler "temel beceriler", okuma-yazma, sayısal ve BİT (Bilgi ve İletişim Teknolojisi) becerileri ise "yaşam becerileri" olarak tanımlanmaktadır (UNESCO 2013).

Hızla gelişen teknolojiler ve değişen dünya bu hıza uyum sağlayabilen, değişebilen insan arayışındadır. Politika yapımcılar ve geleceğin bireylerini yetiştiren eğitimcilerin de bu ihtiyaçları karşılayabilecek öğrenciler yetiştirmeleri gerekmektedir. Bu misyonu kendine görev edinen OECD ilk defa PISA'da 2009 okuma yeterliliğini ölçmek için ek seçenek olarak bilgisayar tabanlı değerlendirme tasarlamış, ardından PISA 2015'e bilgisayar aracılığıyla öğrenci değerlendirmesi yapmıştır. Bu yenilik BİT etki faktörlerinin ortaya çıkmasını sağlamıştır (Xiao vd.

2019). Ayrıca yine güncel konulardan yolla çıkarak OECD küresel yeterliliğe sahip bireyler yetiştirilebilmesi amacıyla (Mostafa 2021) PISA 2018’de küresel yeterlilik alanına yer vermiştir. OECD’e (2018) göre küresel yeterliliğe sahip olan öğrenciler, küresel ve kültürlerarası sorunları inceleyebilir, farklı bireylerle başarılı ve saygılı bir şekilde iletişim kurabilir, sürdürülebilirlik ve kolektif refah için gerekli eylemlerde bulunabilmelidir. Ancak Cobb ve Couch (2021) yapılmaya çalışılan çalışmanın yetersiz olduğunu ve geliştirilmesi gerektiğini ifade etmiştir. Engel vd. (2019) ise potansiyel olarak olumlu sonuçlar vermesine rağmen PISA 2018’de değerler kavramının göz ardı edilmesinden dolayı, küresel yeterliliğin evrensel ve yerel düzeyde ölçülebildiğinden duydukları endişeleri dile getirmişlerdir. Ancak OECD (2018), küresel yeterlilikle test edilen öğrencilerin yine okuma metinleri üzerinden test edildiğini, bu sayede okuma becerisi ile küresel yeterlilik arasındaki ilişkiyi ölçebilmenin mümkün olacağını ifade etmiştir.

Çocuk Vakfı (2006) okuma alışkanlığının kazanılmasında aile, okul ve çevrenin etkili toplumsal kurumlar olduğunu ifade etmiştir. Böylesine çok boyutlu bir süreçte birçok araştırmacı okuma becerisini etkileyen unsurları incelemiştir. Literatürde okuma becerisi üzerine PISA verileriyle yapılan bazı çalışmalara Tablo Ek 1’de yer verilmiştir. Bu kapsamda cinsiyet, sosyoekonomik durum, ebeveynlerin eğitim düzeyleri ve meslekleri, ebeveyn ve öğretmen desteği, BIT kullanımı öğrenciler için önemi bulunan değişkenler arasındadır.

Son yıllarda teknolojinin ilerlemesi okuma alışkanlıklarımızda değişiklik yarattığı gibi bilgisayar uygulamalarının da büyük boyutlu verileri tek seferde ve kısa sürede analiz edilebilmesine olanak sağlamıştır. Bu kapsamda çok boyutlu verilere sahip olan PISA veri setleri de veri madenciliği yöntemlerinin kullanılmasına olanak tanımakta ve başarılı sonuçlar vermektedir. Veri madenciliği (DM) yardımıyla veri depolarında bulunan gizli bilgilere (Kaunang ve Rotikan 2018; Salal vd. 2019) ulaşılabilmektedir. DM yöntemlerinin eğitim verilerine uygulanması Eğitimsel Veri Madenciliği (EDM) olarak tanımlanmaktadır (Asif vd., 2017; Fırat ve Koyuncu, 2020). Jalota ve Agrawal’a (2019) göre EDM, eğitim verilerini ve öğrencileri daha iyi değerlendirebilmek için kullanan bir disiplindir. EDM verilerde yeni kalıplar arayarak yeni algoritmalar veya modeller geliştirirken (Romero ve Ventura, 2020) istatistik, makine öğrenimi ve veri madenciliğinden yararlanır (Mahajan ve Saini, 2020). Bu kapsamda literatürde yapılmış bazı çalışmalar ve kullandıkları yöntemlere Tablo Ek 2’de yer verilmiştir.

Büyük veri kümelerini oluşturan değişkenlerin hepsi ilgilenilen sınıflandırma ile her zaman ilintili değildir. Büyük veri kümeleriyle uğraşmak algoritmayı yavaşlatmasının yanı sıra fazla kaynak tüketimi ve yöntemlerin elverişsiz hale gelmesi gibi dezavantajlara sahiptir. Bu yüzden mümkün olan en iyi sınıflandırma sonuçlarını veren küçük veri setleri tercih edilir (Kursa ve Rudnicki, 2010). Küçük verilerle çalışmak için araştırmacılar tarafından manuel değişken seçimleri gerçekleştirilebilir. Ancak bu seçimler yanlılığa ve zaman kaybına neden olur. Manuel değişken seçimleri yerine literatürde istatistiksel varsayımlara dayanan öznitelik seçim algoritmaları

kullanılması daha avantajlıdır. ML çalışmalarında veri seti içerisinde yararlı değişkenleri (öznitelikleri) seçme ve bulma süreci olarak kabul edilen öznitelik seçim algoritmalarının EDM kapsamında performans ve doğruluklarını karşılaştıran çalışmalar arasında, Zaffar vd. (2017) farklı öğrenci veri setlerini kullandıkları çalışmalarında 6 adet özellik seçim ve 15 adet sınıflandırma algoritması kullandıkları çalışmalarında RF'in daha iyi sonuçlar verdiğini belirtmişlerdir. Jalota ve Agrawal'da (2021) çeşitli özellik (korelasyon ve sarmalayıcı tabanlı) seçim algoritmalarından ve sınıflandırıcılardan (Bayes-Net, JRip, J48, NB, oneR, RF, SMO, Simple Logistic) yararlanarak yaptıkları çalışmalarında SMO ve J48'in korelasyon öznitelik seçim algoritmaları ile en yüksek doğruluk ölçütlerine sahip olduğunu ifade etmişlerdir. Bu çalışmada RF ise en başarılı performansa sahip olmamakla birlikte başarılı sonuçlar elde edilmesini sağlamıştır. Gajwani ve Chakraborty'da (2021) öğrencilerin notlarını doğrudan etkileme olasılığı en yüksek olan özelliklerin bir alt kümesini belirlemek amacıyla farklı algoritmaları (DT, LR, NB, boosting, bagging, voting ve RF) kullandıkları çalışmalarında Boosting, voting ve RF en iyi sonuçları verdiği belirtmişlerdir. Delen'de (2010) çalışmasında okulu bırakma riski olan öğrencilerin yıpranmasına neden olan unsurları belirlemek amacıyla yaptıkları çalışmada çeşitli algoritmalar (ANN, DT, LR, SVM, RF (Bagging), Boosted Trees, Information Fusion (IF)) kullanmıştır. Çalışma sonuçları incelendiğinde doğruluk oranlarının en yüksek olduğu algoritmaların önem sırasına göre IF iken RF (Bagging) olduğu görülmektedir. Ayrıca Han vd. (2019) öğrencilerin problem çözme becerilerini, Son vd. (2020) ise öğrencinin okuma becerisini etkileyen değişkenlerin belirlenmesi için RF algoritmasından yararlanmışlardır. Anand vd. (2021) ise çalışmalarında EDM'de, özellik seçimi olarak Boruta'yı, sınıflandırıcı olarak NB ve DT kullanmışlardır.

Bu çalışmada "Türkiye'deki öğrencilerin okuma becerilerini etkileyen etmenler nelerdir?" sorusuna odaklanılmıştır. Okuma becerisinde yüksek ve düşük puan alan öğrencileri etkileyen önemli değişkenleri tespit edip, ilgili eğitimcilere ve politika yapıcılara yeni ipuçları ve bakış açıları sunabilmek amaçlanmıştır. Bu çalışmanın ilk aşamasında okuma becerisini etkileyen değişkenlerin belirlenmesi sırasında RF temeline dayanan Boruta algoritması kullanılmıştır. İkinci adımda ise belirlenen bu değişkenleri oluşturan sorular kullanılarak yine Boruta algoritması ile ön plana çıkan sorular belirlenmiştir. Petko vd. (2017) PISA anket maddelerinin genel ve öğrenci beyanlarına dayandırıldığı için bulguların olası yorumlarının dikkatlice ele alınmasından bahsetmişlerdir. Bu kapsamda çalışmanın üçüncü aşamasında ise elde edilen sorular /değişkenlere ait olasılık tabloları oluşturulmuştur. Çalışmanın son kısmında, elde edilen bulgular değerlendirilerek PISA 2018 Türkiye örnekleminde elde edilen çıktılar doğrultusunda öğrencilerin okuma becerilerinde öne çıkan değişkenler üzerine tartışılmıştır.

Yöntem

Araştırma Modeli

Geniş ölçekli değerlendirmelerde öğrencinin aldığı puan değil örneklemin aldığı puan dikkate alınmaktadır (Pejic ve Molcer, 2019). Bu puanlar makul değerler

(plausible values-PVs) olarak ifade edilir. Arıkan vd. (2020) geniş ölçekli test (PISA, PIAAC ve TIMSS) verilerinin örneklem ağırlıklarını dikkate alan yöntem ve yazılımlar kullanılması gerektiğinden, Tat vd. (2019) ise geniş ölçekli uygulamalarda verilerin (öğrenci, okul, bölge, ülke, vb.) hiyerarşik bir yapıya sahip oldukları için makul değerlerin öğrencilerin bireysel puanlarını temsil etmeyeceğini belirtmişlerdir. Özkan (2021) ise “PISA makul değerlerinde öğrencilerin bireysel değerleri olarak kullanılmaması” gerektiğini belirtmiş, ancak PISA “kendisi tarafından belirlenen düzeyleri kullanarak ülkedeki öğrencileri PV değerlerini dikkate alarak (cinsiyetler, okullar, bölgeler bazında sınıflandırdığı) “değerlendirmeler yaptığını ifade etmiştir. Yılmaz vd. (2020) tarafından hazırlanan PISA ve Türkiye (2000 - 2018) raporunda ise “PISA beceri ölçeğinde belirli bir noktaya yerleştirilen bir öğrenci o düzey ve o düzeyin altındaki görevleri yerine getirebilir” ifadesini kullanmışlardır. Öğrencilere ait PV değerlerinin bireysel olarak kullanılamayacağı ancak bu değerlere dayanarak belli bir düzeyde yer almaları durumunda değerlendirmeye tabi tutuldukları PISA raporlarında karşımıza çıkmaktadır. PISA'nın performans yordayıcı çalışmalarında öğrencilere ait doğrudan kullanılacak net bir başarı puanı olmadığı için EDM çalışmalarında öğrenci hakkında genel bir bilgiyi temsil etmesi bakımından bu tür çalışmalarda makul değerler kullanılmaktadır (Bezek Güre vd., 2020; Depren ve Depren, 2021; Dong ve Hu, 2019; Gamazo ve Martínez-Abad, 2020; Koyuncu ve Gelbal, 2020; Özkan, 2021; Pejic ve Stanic Molcer, 2018; Son vd., 2020). Bu çalışma bir sınıflandırma temelli olmasından dolayı yine direk öğrencinin bireysel puanı üzerine odaklanarak analizler yapılmadığı, öğrencinin temsil ettiği sınıf dikkate alınarak değerlendirildiğinden dolayı öğrencilere ait PV değerleri kullanılması uygun bulunmuştur.

Veri Toplama Araçları

Bu araştırmada veri toplama aracı olarak, PISA 2018 Türkiye örnekleme ait öğrenci anketinden(OECD 2019c) yararlanılmış nicel ilişkisel bir çalışmadır. Öğrencilerin okuma beceri testinden alınan puanlar (PV1-PV10) literatürde bazı çalışmalarda (Bezek Güre vd. 2020; Dong ve Hu 2019; Lee 2018; Pejic ve Stanic Molcer 2018) ve PISA'nın sonuç raporlarında kullandığı gibi ortalamaları alınmıştır. Erişime açık bir veri seti olmasından dolayı verilerin kullanımı için özel izin talep edilmemiştir.

Etik Kurul Onayı

Çalışmada OECD (2019b) tarafından tüm araştırmacıların kullanımına sunulan veri seti kullanılmıştır. Bu nedenle çalışma için herhangi bir etik bildirim ihtiyacı duyulmamıştır.

Verilerin analizi

Öznitelik seçimi (Feature Selection)

Öznitelik seçimi (FS), örüntü tanımda yaygın olarak kullanılan bir tekniktir. Verinin içinde bulunan hatalı (gürültülü) ve gereksiz özelliklerin (değişkenlerin) çıkarılması ile verinin daha anlaşılabilir olmasını (Venkata ve Lingamgunta, 2020;

Yan ve Zhang, 2015) sağlayarak veri depolama gereksinimlerini, eğitim ve uygulama sürelerinin azalmasını (Guyon ve Elisseeff, 2003), algoritmaların hızını artırmayı ve tahmin doğruluğu için performans iyileştirmeyi sağlar (Guyon ve Elisseeff, 2006).

Rastgele Ormanlar (Random Forest)

Parametrik olmayan istatistiksel yöntemler arasında yerini alan RF algoritması, Breiman (2001) tarafından geliştirilmiştir. RF algoritması, sınıflandırma, öznitelik seçimi ve regresyon konularında kullanılabilir (Lahouar ve Slama, 2015). Topluluk öğrenme yöntemi olan RF, bir dizi karar ağaçlarının birleşmesi ile daha iyi bir performansa sahip daha büyük bir ağaç oluşturarak (Unpingco 2019) kümeleme ve bootstrap fikirlerini birleştiren; regresyon ve sınıflama problemlerinde kullanılan güçlü bir yöntemdir (Bezek Güre vd., 2020).

Boruta Algoritması

RF sınıflandırma algoritması etrafında oluşturulmuş bir sarmalayıcı yöntemdir (Ahmed vd., 2021; Kurşa ve Rudnicki, 2010). Bu algoritma genellikle veri kümesini çoğaltarak veri setini genişletir ve rastgele ormandan elde ettiği Z-skorumu kullanır (Iman ve Ahmad, 2020). Gölge öznitelikler arasından en yüksek Z-skoru hangi özniteliklerin önemli/önemsiz olduğunun belirlenmesinde referans olarak kullanılır (Kurşa ve Rudnicki, 2010). Referansın üzerindeki öznitelikler önemli, düşük olanlar önemsiz olarak kabul edilir. Önemsiz değişkenlerin çıkarılması ile analiz tekrar edilir. Sonuçta önemli öznitelikler kalana kadar analiz tekrarlanır.

EDM kapsamında sınıflandırma temelli çalışmalarda farkı özellikler dikkate alınarak sınıflandırmalar yapılmaktadır. Asif vd. (2017) öğrenci performansları başarılı ve başarısız gruplara ayırırken Kasap vd. (2021) ise 6 düzeye ayrılan PISA okuma beceri puanlarının ilk üç düzeyini düşük, diğer üç düzeyini ise yüksek olarak kodlamıştır. Emdadi ve Eslahchi (2021) ve Xu vd. (2019) çalışmalarında bağımlı değişkenin medyan değerini baz alarak sınıflandırma yapmışlardır. İlk olarak 6890 öğrencinin okuma becerileri puan (PV1-PV10) ortalamaları (M_{PVREAD}) hesaplanmıştır. Aritmetik ortalamadan duyarlı bir ortalamasından dolayı çalışmanın daha kolay ifade edilmesini ve sınıflandırma performansını artırmak amacıyla bağımlı değişken olan M_{PVREAD} Asif vd. (2017), Mahajan ve Saini (2020), Son vd. (2020) ve Han vd. (2020) çalışmalarında kullandıkları gibi sınıflandırılmasına karar verilmiştir. M_{PVREAD} ' in medyan değeri ($Med_{PVREAD}=468$) hesaplanarak medyan değerinin altında kalan sonuçlara 1 (düşük), medyan değerine eşit ve üzerinde değer alan ortalamalar ise 2 (yüksek) olacak şekilde yeniden kodlanmıştır (PVREAD).

PISA, testlerde kullandıkları değişkenleri; basit indisler, ölçek indisleri ve trend ölçeği indisleri olmak üzere üç başlık altında toplamıştır. Basit indisler, aritmetik dönüşüm ve yeniden kodlanarak elde edilmiştir. Ölçek indisleri ise birden fazla maddenin ölçeklendirilmesi ile elde edilen parametrelerden oluşturulmuştur (OECD 2019b). Çalışmada kullanılan indislere Tablo 1'de yer verilmiştir.

Tablo 1.*Çalışmada Kullanılan İndis Listesi*

Öğrenci İndisleri
ADAPTIVITY, AGE, ATTIMM, ATTLNACT, AWACOM, BEINGBULLIED, BELONG, BFMJ2,BMMJ1, BSMJ, COBN_F, COBN_M, COBN_S, COGFLEX, COMPETE, CULTPOSS, DIRINS, DISCLIMA, DISCRIM, DURECEC, EMOSUPS, ESCS, EUDMO, FISCED, FISCED_D, GCAWARE, GCSELFEFF, GFOFAIL, GLOBMIND, GRADE, HEDRES, HISCED, HISCED_D, HISEI, HOMEPOS, , ICTRES, IMMIG, INTCULT, ISCED, ISCEDD, ISCEDL, ISCEDO, JOYREAD, LANGN, LMIN, MASTGOAL, METASPAM, METASUM, MISCED, MISCED_D, MMIN, PARED, PAREDINT, PERCOMP, PERCOOP, PERFEED, PERSPECT, PISADIFF, REPEAT, RESILIENCE, RESPECT, SCREADCOMP, SCREADDIFF, ,SMINS, ,STIMREAD, STRATUM, SWBP, TEACHINT, TEACHSUP, TMIN, UNDREM, WEALTH, WORKMAST
BİT İndisleri
ICTHOME, ICTSCH, ENTUSE, USESCH, AUTICT, INTICT, COMPICT, SOIAICT, ICTOUTSIDE, ICTCLASS, HOMESCH

Veri setindeki yüksek miktarda eksik veya tanımsız bilgiler içeren indisler silinmiş (ATTIMMP, BODYIMA, CHANGE, CNTSCHID, CNTSTUID, CURSUPP, EMOSUPP, EFFORT1, EFFORT2, EMOSUPS, FCFMLRTY, FLCONFIN, FLCONICT, FLFAMILY, FLSCHOOL, GCAWAREP, INFOCAR, INFOJOB1, INFOJOB2, INTCULTP, JOYREADP, LANGFATHER, LANGFRIEND, LANGMOTHER, LANGSCHMATES, LANGSIBLINGS, LANGTEST_COG, LANGTEST_PAQ, LANGTEST_QQQ, OCOD1, OCOD2, OCOD3, OECD, PASCHPOL, PQSCHOOL, PRESUPP, PROG, SCCHANGE, SOCONPA, STUBMI, SUBNATIO) ,Cinsiyet (Gender) değişkeni de dahil edilerek, biri bağımlı (PVREAD) olmak üzere toplamda 86 değişken analize dahil edilmiştir. Belirlenen bu 86 değişken içinde eksik veriler için interpolasyon yöntemi ve analiz için R programına ait randomForest (Liaw ve Wiener, 2018) ve Boruta (Kursa 2020) paketleri kullanılmıştır. Tablo 1’de bu çalışmanın 1. aşamasında kullanılan 84 indis ismine yer verilmiştir. Her iki aşamada da Z dönüşüm (meanImp) değerleri dikkate alınarak değerlendirme yapılmıştır. Uygulamanın ilk aşamasında 100 ağaç ve accuracy (doğruluk) değeri 69,7 ile okuma becerilerini (PVREAD) etkileyen indisler içinden belirgin 27 değişken (indis) elde edilmiştir.

Çalışmanın 2. aşamasında, 27 indise ait sorular belirlenmiş ve analize dahil edilmiştir. BSMJ (Mesleki Beklenti) indisine ait sorular elde edilemediğinden, analize ilave olarak bu indiste eklenmiştir. Belirlenen 159 değişken içinden analiz sonuçları sonunda 100 ağaç ve accuracy değeri 67,2 ile 52 değişkenin (sorunun) ön plana çıktığı görülmüştür.

Bulgular

Uygulamanın ilk aşamasında okuma becerilerini etkileyeceği düşünülen 85 değişkene ait ilk analiz sonuçları Şekil 1’de görüldüğü gibidir. Mavi kutu grafikleri, bir gölge özneliğinin minimum, ortalama ve maksimum Z puanına karşılık gelir. Kırmızı, sarı ve yeşil kutu grafikleri sırasıyla reddedilen, belirsiz ve onaylanan niteliklerin Z puanlarını temsil etmektedir. Önemsiz ve belirsiz değişkenlerin çıkarılması ile analiz tekrarlanmış ve çalışmada kullanılacak 27 değişken (indis) elde edilmiştir ve bu değişkenler Tablo 2’de önem sıralarına göre listelenmiştir.

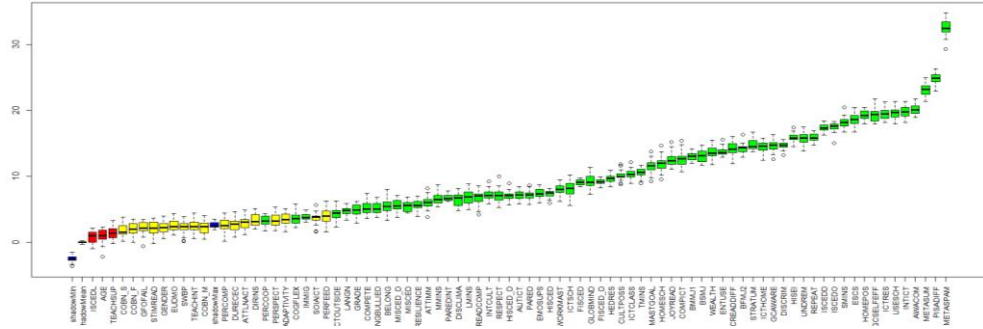
Tablo 2.
Okuma Becerisinde Ön Plana Çıkan İndisler

indisler	meanİm p	indisler	meanİm p	indisler	meanİm p
1.METASPA M	47,77	10.HOMEPOS	26,03	19.SCREADDIF F	17,24
2.PISADIFF	33,15	11.SMINS	24,30	20.GCAWARE	17,23
3.METASUM	30,72	12.GCSELFEF F	24,26	21.BSMJ	17,16
4.ESCS	30,56	13.ICTHOME	20,62	22.COMPICT	16,94
5.USESCH	27,46	14.STRATUM	19,38	23.JOYREAD	14,28
6.ISCEDO	27,33	15.UNDREM	19,29	24.FISCED	13,44
7.INTICT	26,66	16.WEALTH	18,35	25.TMINS	13,33
8.ICTRES	26,53	17.ENTUSE	17,97	26.GLOBMIND	10,35
9.AWACOM	26,04	18.HOMESCH	17,65	27.MISCED	8,43

Tablo 2 incelendiğinde, indislerin BİT kaynaklarına ulaşım ve kullanımı (USESCH, ENTUSE, HOMESCH, ICTHOME), BİT’e ilgi ve algılanan yeterlilik (INTICT, COMPICT), beklenen mesleki statü (BSMJ), ekonomik sosyal kültürel düzey (ESCS, HOMEPOS, ICTRES, WEALTH, FISCED, MISCED), eğitime ayrılan süre (SMINS, TMINS), üstbilgi (METASPAM, METASUM, UNDREM), okumaya yönelik tutumlar (JOYRED, PISADIFF, SCREADDIFF), dünya ve diğer kültürlerle duyulan merak (GLOBMIND, AWACOM, GCSELFEEF, GCAWARE) ve öğrencinin okuduğu okul türü (STRATUM, ISCEDO) olmak üzere 9 başlık altında gruplandırıldığı görülmüştür.

Çalışmanın 2. aşamasında 9 başlık altında gruplanan indislere ait 159 sorunun analizi sonucunda (Şekil 2) elde edilen önemsiz değişkenler çıkarılmış elde kalan 52 soru Tablo 3’te indis isimleri ve önem sıralarına göre verilmiştir. Öne çıkan indislerin (STRATUM hariç) okuma beceri puan ortalamaları (M_{PVREAD}) ile arasında istatistiksel olarak anlamlı ilişkiler olduğu (Tablo 4) en yüksek ilişkilerin 1.aşamada ön plana çıkan indislerden (Tablo 2) METASPAM, METASUM ve PISADIFF olduğu tespit edilmiştir.

Şekil 1.
Birinci Aşama Analiz Sonuçları



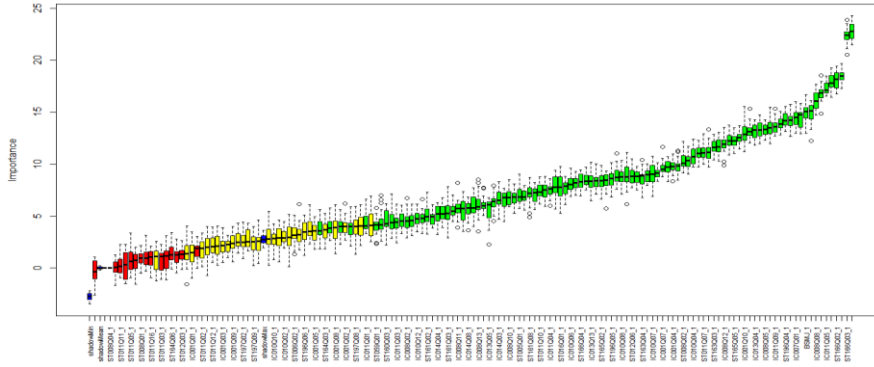
Tablo 3.
Okuma Becerisinde Ön Plana Çıkan Sorular Ve İndisleri

Değişken	Indis ismi	meanImp	Değişken	Indis ismi	meanImp
ST013Q01	HOMEPOS	31,19	ST011Q07	HOMEPOS	13,60
ST166Q03	METASPAM	29,63	ST007Q01	FISCED	13,48
ST166Q02	METASPAM	24,18	IC008Q04	ENTUSE	13,36
ST059Q03	SMINS	23,60	IC010Q04	HOMESCH	13,29
IC008Q08	ENTUSE	22,74	IC001Q06	ICTHOME	13,13
ST163Q02	PISADIFF	22,40	ST164Q05	UNDREM	11,94
ST002Q01	STRATUM	22,24	IC008Q07	ENTUSE	11,80
IC011Q05	USESCH	22,13	IC008Q09	ENTUSE	11,68
BSMJ	BSMJ	18,29	ST166Q04	METASPAM	11,65
ST060Q01	TMINS	18,14	ST165Q02	METASUM	11,60
ST166Q05	METASPAM	18,07	ST160Q04	JOYREAD	11,38
IC001Q10	ICTHOME	18,04	IC010Q07	HOMESCH	11,00
ST196Q04	GCELFEFF	18,04	IC013Q01	INTICT	10,90
IC008Q05	ENTUSE	17,54	IC013Q13	INTICT	10,87
ST012Q05	HOMEPOS	17,39	ST163Q04	PISADIFF	10,33
IC014Q06	COMPIC	17,27	ST005Q01	MISCED	9,93
IC013Q04	INTICT	17,13	ST161Q08	SCREADDIFF	8,24
ST197Q01	GCAWARE	17,03	ST161Q07	SCREADDIFF	7,94
IC011Q09	USESCH	16,80	ST166Q01	METASPAM	7,24
IC011Q06	USESCH	16,79	ST164Q04	UNDREM	6,78
ST165Q04	METASUM	16,03	ST164Q01	UNDREM	6,61
IC011Q08	USESCH	15,59	ST165Q03	METASUM	5,43
IC008Q12	ENTUSE	15,28	ST164Q03	UNDREM	5,28
ST165Q05	METASUM	14,90	ST164Q02	UNDREM	4,67
ST163Q03	PISADIFF	14,06	ST165Q01	METASUM	4,48

ST196Q02 GCSELFEFF 14,03 ST164Q06 UNDREM 2,99

Şekil 2.

İkinci Aşama Analiz Sonuçları



Tablo 4.

M_{PVREAD} ve İndis Korelasyon Tablosu

	METASPAM	METASUM	PISADIFF	HOMEPOS	UNDREM
<i>M_{PVREAD}</i>	,423**	,371**	-,356**	,337**	,294**
	GCSELFEFF	BSMJ	SCREADDIFF	JOYREAD	SMINS
<i>M_{PVREAD}</i>	,269**	,259**	-,252**	,237**	,231**
	COMPACT	FISCED	USESCH	INTICT	GCAWARE
<i>M_{PVREAD}</i>	,206**	,198**	-,192**	,191**	,187**
	MISCED	ENTUSE	ICTHOME	HOMESCH	TMINS
<i>M_{PVREAD}</i>	,187**	,124**	,115**	-,064**	,041**

** . Correlation is significant at the 0.01 level (2-tailed).

Tablo 5.

Okuma Becerilerine Ait Korelasyon Tablosu

	<i>M_{PVMATH}</i>	<i>M_{PVSCI}</i>
<i>M_{PVREAD}</i>	,872**	,922**
<i>M_{PVMATH}</i>	1	,918**

** . Correlation is significant at the 0.01 level (2-tailed).

Okumaya yönelik tutumlar

1. aşamada belirlenen okumaya yönelik tutumlar başlığında yer alan okumayı sevme (JOYRED), zorluk algısı (PISADIFF) ve güçlük algısı (SCREADDIFF) indislerinin öğrencilerin okuma becerileri üzerine etkili oldukları bulunmuştur. Bu üç indis arasında PISADIFF'in önem düzeyinin altıncı sırada olduğu Tablo 2'de görülmüştür. Ayrıca SCREADDIFF ve PISADIFF'in PVREAD'le aralarında

istatistiksel olarak anlamlı ancak negatif yönlü bir ilişki olduğu tespit edilmiştir (Tablo 4).

Üstbiliş

Öğrencilerin okumayla ilgili bilişsel durumlarını değerlendirmek üzere Anlamak ve hatırlamak (UNDREM), Özetlemek (METASUM) ve Güvenilirliği değerlendirmek (METASPAM) indislerinin PVREAD ile aralarında istatistiksel olarak anlamlı ilişkiler olduğu tespit edilmiştir (Tablo 4).

Okuma becerisine göre sınıflandırılmada METASPAM ve METASUM öncelik sıralamasında öne çıkan indisler arasındadır (Tablo 2). Analiz sonuçlarına bakıldığında okuma stratejilerinin (METASPAM, METASUM ve UNDREM) okuma becerilerinin sınıflandırılmasında önemli bir rol oynadıkları doğrulanmıştır.

Eğitime ayrılan süre

Öğrencilerin fen bilimleri (SMINS) ve toplam derslere (TMINS) ait aldıkları haftalık ders saatlerinin okuma becerisi üzerinde istatistiksel olarak anlamlı bir ilişkiye sahip olduğu görülmekle birlikte (Tablo 4) TMINS'e ait korelasyon ($r=0.041$) değerinin neredeyse sıfır olduğu tespit edilmiştir.

Okuma becerisi ile fen bilgisi arasındaki ilişkiyi gözlemleyebilmek amacıyla, okuma (MPVREAD), matematik (MPVMATH) ve fen bilgisi (MPVSCI) beceri puan ortalamaları arasındaki ilişkiler incelenmiştir (Tablo 5). En yüksek korelasyon değerinin okuma becerisi (MPVREAD) ortalaması ile fen bilgisi beceri ortalaması arasında (0.92) olduğu tespit edilmiştir.

Dünya ve diğer kültürler hakkında bilgi

Çalışmanın 1. aşamasında dünya ve diğer kültürlere duyulan merak başlığı altında (GLOBMIND, AWACOM, GCSELFEEFF, GCAWARE) dört indisin ön plana çıktığı görülmektedir (Tablo 2). Ancak çalışmanın 2. aşamasında bu indisler içinde sadece iki tanesi ön plana çıkabilmiştir. GCSELFEEFF, bir öğrencinin küresel konuları açıklamaya veya tartışmaya ilişkin öz yeterliliğini ele alırken, GCAWARE ise bir öğrencinin küresel sorunlara ilişkin farkındalığını sorgular. Bu iki indisinde öğrencilerin okuma becerileri ile pozitif yönde anlamlı ilişkiye sahip oldukları tespit edilmiştir (Tablo 3).

Ekonomik-Sosyal-Kültürel Düzey

Çalışmanın 1. aşamasında ekonomik-sosyal-kültürel düzey başlığı altında (ESCS, HOMEPOS, ICTRES, WEALTH) dört indis belirlenmiştir (Tablo 2). ESCS üç indisten oluşan bileşik bir puandır. Ebeveyn mesleki durumu (HISEI), ebeveyn eğitimi (PAREDINT) ve ev halkına ait ev eşyaları (HOMEPOS) endekslerinden oluşur. HOMEPOS indisi, öğrencilerin evlerinde bulunan eşyalarını, evdeki mevcut kitap sayısı da dahil olmak üzere zenginlik (WEALTH), kültürel varlıklar (CULPOSS) ve evde eğitim kaynakları (HEDRES) endekslerinden elde edilir. ICTRES indisi ise, sahip olunan BİT kaynaklarından oluşturulmuştur.

Çalışmanın 2. aşamasında bu indisler içinde HOMEPOS, FISCED ve MISCED olmak üzere bu üç indisi temsil eden sorular ön plana çıkmıştır. FISCED ve MISCED indisleri PAREDINT indisinin ebeveynlere ait olan bilgilerini içermektedir. Ayrıca bu üç indis ile okuma becerisi arasında istatistiksel olarak anlamlı ilişkiler tespit edilmiştir (Tablo 4).

BİT kaynaklarına ulaşım ve kullanım

Tablo 2'de BİT kaynaklarına ulaşım ve kullanım başlığı altında yer alan USESCH okulda kullanılan, ENTUSE okul dışı etkinliklerde kullanılan, HOMESCH ise okul dışında ödevlerde kullanılan dijital cihaz kullanım sıklığını temsil eder. ICTHOME ise evde kullanılan BİT cihazlarından elde edilmiş indislerdir.

Çalışmanın 2. aşamasında yine bu dört indisi temsil eden sorular ön plana çıkmış (Tablo 3), okuma becerilerinin sınıflandırılmasında önemli bir rol oynadıkları doğrulanmıştır. Ayrıca bu dört indis ile okuma becerisi arasında istatistiksel olarak anlamlı ilişkiler tespit edilmiştir (Tablo 4). Ancak USESCH ve HOMESCH indislerinin PVREAD ile istatistiksel olarak anlamlı ve negatif bir ilişkiye sahip iken, HOMESCH korelasyon değerinin ($r=-0.064$) çok düşük olduğu gözlenmiştir.

BİT'e ilgi ve algılanan yeterlilik

Tablo 2'de BİT'e ilgi ve algılanan yeterlilik başlığı altında toplanan (INTICT, COMPICT) iki indis, çalışmanın 2. aşamasında önemlerini korumuşlardır. Bu iki indiste öğrencilerin dijital medya ve dijital cihazlara olan ilgilerini (INTICT) ve öğrencilerin kendilerini ne kadar yeterli (COMPICT) bulduklarını değerlendirmeye çalışmaktadır. Her iki indisinde okuma beceri ortalamaları ile pozitif ve istatistiksel olarak anlamlı ilişkiye sahip olduğu tespit edilmiştir (Tablo 4).

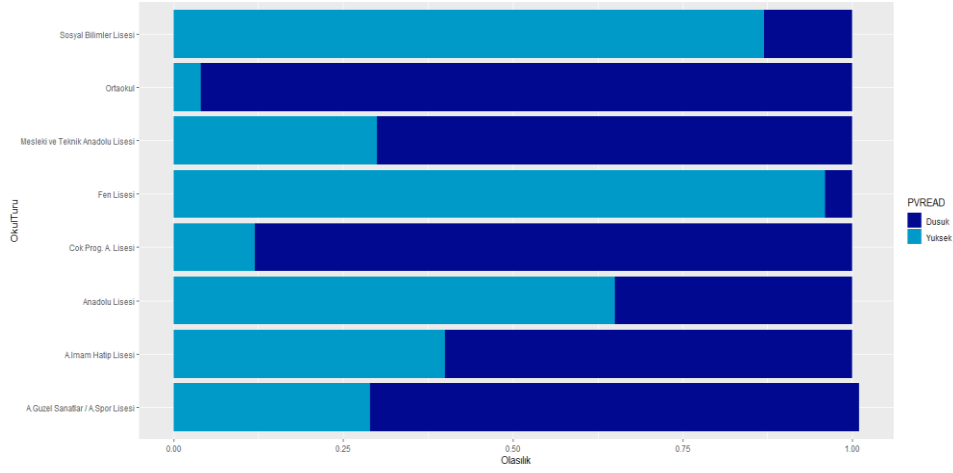
Mesleki Beklenti

Çalışmanın iki aşamasında da kullanılan BSMJ indisi, Uluslararası Çalışma Örgütü (ILO)'nun yapmış olduğu sınıflandırma yapısına göre otuz yaşlarında öğrencilerin kendilerini hangi işte çalışıyor olduklarını umdukları sorularına verilen yanıtların düzenlenmiş halidir. Alınan yüksek puanlar öğrencinin beklentisinin yüksek mesleki statü hedeflediğini göstermektedir (OECD 2020). BSMJ ile okuma becerisi arasında pozitif yönde istatistiksel olarak anlamlı bir ilişki tespit edilmiştir (Tablo 4).

Okul türü

Öğrencilerin okudukları okul türü ya da bilgisi indisi olarak tanımlanan STRATUM, çalışmanın iki aşamasında da önemlilik arz etmiştir (Tablo 2, Tablo 3). Öğrencilerin gittiği okul türlerine göre yeniden düzenlendikten sonra okuma beceri puanlarına göre sınıflandırılmıştır. Olasılıklar dikkate alınarak hazırlanan Şekil 3 aşağıdaki gibi elde edilmiştir. Okuma becerisi yüksek öğrencilerin Anadolu Lisesi, Fen ve Sosyal Bilimler Lisesindeki öğrenciler olduğu tespit edilmiştir.

Şekil 3.
Okul Türlerinin Okuma Beceri Durumları



Tartışma, Sonuç ve Öneriler

Francis Bacon'un 1597 tarihinde yazdığı "bilginin kendisi güçtür" sözü günümüzde teknolojinin hızla ilerlediği bu çağda daha belirgin olarak hissedilmektedir. Okuma, bilgiyi besleyen unsurların başında gelir. Okuma becerisi bu yüzyılın hızına yetişebilmemiz için eskisinden de önemli bir yere sahiptir. Bugünü modern ve karmaşık olarak tanımlayan Fabunmi ve Folorunso'a (2010) göre daha iyi yaşamak ve başarılı olabilmek için okuma becerilerine ihtiyaç duymaktayız. Nzeyimana ve Bazimaziki'e (2020) göre günümüz dünyasında teknolojinin hızlı ilerlemesi sonucunda temelsiz bilgilere güvenmek ve tam olarak bilgilenmemekle sonuçlanacak başarısız okuma becerileri ile başa çıkabilmek için yeni stratejilerin belirlenmesi gerekmektedir.

Bireyin üstbilis stratejilerini etkin şekilde kullanmasının, öğrenme kalitesi üzerinde önemli bir etkisi olduğu kabul edilir (Arslan 2020). Üstbilis farkındalığı olan öğrenciler hangi stratejileri nerede ve ne zaman kullanacaklarını bilirler (Fırat ve Koyuncu, 2020). Üstbilis indislerinin okuma becerilerinde etkili olduğu Depren ve Depren (2021); Fırat ve Koyuncu (2021) ve Son vd. (2020) çalışmasında olduğu gibi bizim çalışmamız tarafından da desteklenmiştir.

Dijital dünyanın hızlı yükselişi kullanıcıların daha dikkatli okumalar yapmalarını zorunlu hale getirmiştir. UNESCO'da (2013) ifade ettiği gibi okuryazarlık tanımı artık yaşam becerilerinden oluşmaktadır. Bugün dünyanın birçok yerinde dijital okuma becerisi, bireyin hedeflerine ulaşma ve topluma katılma becerisinin anahtarı (OECD 2016) olarak görülmektedir. Teknolojinin çok hızlı ilerlemesi ile öğrencilerin karşılaştıkları sorunlarla başa çıkabilmeleri için eğitim

sistemlerinin bu değişimleri doğrudan sınıf eğitimlerine taşınması (Henry 2006; Keskin 2014), öğrencilerin dijital okuma beceri eğitimleri ile desteklenmesi gerekmektedir.

Günümüzde teknoloji ve internetin bizlere sunduğu sınırsız bilgiye erişimin sonunda hala araştırma konuları hakkında bir fikir sahibi olmamak, yetersiz dijital medya okuryazarlığına işaret edebilir (OECD 2018). Küresel yeterliği gelişmiş öğrencilerin disiplinlere (Fen, Edebiyat, Matematik, Biyoloji, Tarih vb.) küresel bir bakış açısı geliştirmesi, derinlemesine öğrenmesi ve eğitim hayatlarında daha başarılı olmalarını sağlayabilir (Tedmem 2019). Ayrıca hem düşük küresel yeterliğe sahip genç kuşağın, taraflı ve sahte haberler tarafından kandırılabilir (OECD 2018). Bu kapsamda okul yöneticileri ve öğretmenlerin küresel yeterliliği geliştirmek adına öğrencilerin dijital olanakları daha anlamlı ve üretken şekilde kullanmaları sağlanmalı, üstbiliş stratejileri ile desteklenmeleri, amaçlanan sürdürülebilirlik ve kolektif refah fikirlerine bir adım daha yaklaşılmasına olanak sağlayacağı düşünülebilir. Okuma becerisi ile küresel yeterlilik arasındaki ilişkiyi ölçebilmenin mümkün olacağını ifade eden OECD'nin (2018) görüşü çalışmamızla paralellik göstermiştir. Kültürel birikim ve küresel sorunlara karşı ilgisi olan öğrencilerin okuma becerilerinin yüksek olduğu belirlenmiştir.

Oriogu vd. (2017) çalışmalarında öğrencilerin okuma alışkanlığını engelleyen en önemli faktörün sosyal medya (Facebook, Whatsapp, Twitter vb.) olduğunu ifade etmiştir. UNESCO (West ve Chew, 2014) raporuna göre gelişmekte olan beş ülkenin üçte ikisi, okumalarını telefonlarından yapmaktadır (OECD, 2016; West ve Chew, 2014). Yine aynı rapora göre mobil okuyucular olarak tanımlanan bu kişilerin çoğunun genç ve normale göre daha fazla eğitilmiş oldukları ifade edilmiştir. Okuma alışkanlığının şekil değiştirdiği günümüzde e-okumaların avantajları gençleri elektronik içeriklere yönlendirmektedir (Mushtaq vd., 2020). Adeyokun vd. (2020) ortaokul öğrencileri üzerine yaptıkları çalışmada akıllı telefonlar, dizüstü bilgisayarlar aracılığı ile internet üzerinden yapılan okumaların öğrencilerin okuma alışkanlıklarını olumlu yönde etkilediğini ifade etmişlerdir. Srijamdee ve Pholphirul(2020) ve Vázquez-Cano vd. (2020) eğitimleri sırasında BİT kullanan öğrencilerin olumlu etkilere sahip olduklarını, ancak eğitim dışı aktivitelerde BİT kullanımının akademik başarılar üzerinde bir etkisinin olmadığını ifade etmişler. Petko vd. (2017) ise evde dijital eğlence amaçlı, Xiao vd. (2019) ise sosyal etkileşim için BİT kullanan öğrencilerin akademik başarıları ile arasında negatif bir ilişki olduğunu belirtmişlerdir. Gubbels vd.(2020) ve Xiao vd. (2019) BİT kullanımı ile öğrenci başarıları arasındaki ilişkinin ters U şeklinde olduğu, aşırı kullanımın pozitif olan etkiyi tersine çevirebileceğini belirtmişlerdir. Öğrencilerin internette fazla zaman geçirmelerinden çok, zamanı nasıl geçirdikleri daha önemlidir. Henry'nin (2006) ifade ettiği gibi öğrencilerin hayata hazırlanmaları için öğretmenlerinin sınıflarında gelişen yeni okuryazarlıkları anlamaları ve çözüm yolları ararken öğrencileri internetten uzaklaştırmamaları gerekmektedir. Ayrıca öğrencilerin teknolojinin eleştirel ve nitelikli kullanımı konusunda bilgilendirilmeleri, yaşlarına uygun olmayan içeriklere karşı kaçınma stratejileri öğretilmelidir (Vázquez-Cano vd. 2020). Lee ve

Wu (2012), öğrencilerin BİT algıları ile akademik performansları, BİT tutumları ile öğrencilerin okuma performansları (Ertem, 2021; Petko vd., 2017) arasında pozitif ilişkiler tespit etmişlerdir. Benzer şekilde Hu vd. 'de (2018) BİT'e ilgi ve yeterlilik bildiren, Srijamdee ve Pholphirul (2020) ise BİT'e karşı deneyim ve aşına olan öğrencilerin okuma, fen ve matematik alanlarında yüksek performansa sahip olduklarını ifade etmişlerdir. Bizim sonuçlarımız da literatürle uyumludur. BİT'e karşı ilgi ve yeterlilik bildiren öğrenciler yüksek okuma beceri puanlarına sahiptir. Teknolojinin hızla ilerlediği bu dönemde, evlerimizde kullanmaya başladığımız akıllı cihazları da düşünecek olursak, öğrencilerin BİT'lere yatkınlıklarının artırılması ve kendilerini bu konuda yeterli hissettirilmeleri önem arz etmektedir.

Öğrencilerin kariyer beklentileri, gelecekteki istihdam fırsatları ile eğitim planlarının uyumlu olup olmadığının ortaya konması açısından önemlidir (Sikora ve Pokropek, 2006). Consulting ve Trust (2013) komisyon toplantısında, 18-24 yaşındaki genç neslin tam zamanlı eğitim ve istihdamda olmadıklarını, bu genç işsizliğin düşük okuryazarlıkla ilişkili olduğunu belirtmişlerdir. Gençlerin gelecek taleplerini yükselterek okuryazarlığını artırmasına odaklanılması gerektiği, bunun içinde 11-14 yaş grubunun özel odak noktası olarak belirlenmesi, gerçekçi istihdam taleplerini oluşturabilmeleri için özellikle orta öğretimin ilk yıllarında desteklenmeleri gerektiği ifade edilmiştir. Gamazo ve Martinez-Abad (2020) çalışmalarında belirttikleri gibi beklenen mesleki statü öğrencilerin performansını artıran bir unsurdur. Bizim çalışmamızda da mesleki beklentisi yüksek olan öğrencilerin okuma beceri puanlarının yüksek olduğunu tespit edilmiştir.

Okuma becerisini etkileyen değişkenler arasında okul türü de (Çelik ve Yurdakul, 2020; Ertem, 2021; Fırat ve Koyuncu, 2020; OECD, 2019a) önemli etkenlerden biri olarak karşımıza çıkmıştır. Eğitim uygulamaları ve müfredatlar öğrencinin akademik başarısına katkıda bulunabilir (Flores-Mendoza vd., 2021). Suna vd. (2020) okullar arası ve okul içi başarı performanslarının Türkiye'nin uzun zamandır tartışılan konularından biri olduğunu, bu farklılığın ortaöğretimde başlayıp artarak devam ettiğini ifade etmiştir. Okuma becerisi bakımından daha önce yapılan PISA sonuçları içinde de en yüksek puan ortalaması yine fen liselerine aittir (Yılmaz vd., 2020). Okul türleri arasındaki farklılığı azaltmak için politika yapıcılar ve eğitimciler tarafından düşük başarı gösteren okul türlerinin desteklenmesi (Suna vd., 2020) ve müfredatlarının tekrar gözden geçirilmesi uygun olacaktır.

Ebeveynlerin sosyoekonomik geçmişi öğrencinin akademik performansını etkileyen önemli faktörler arasında kabul edilmektedir (Fırat ve Koyuncu, 2020; Geske ve Ozola, 2008; Hu vd., 2021).PISA 2018'de Son vd. (2020) Koreli öğrencilerin sosyoekonomik göstergeleri düşük olan öğrencilerin okuma beceri puanlarının düşük olduğu tespit etmiştir. Chung vd. (2021) Finlandiya ve Koreli öğrencilerde sosyoekonomik göstergeler ile okuma becerileri arasında olumlu yönde bir etki bulurken Suna vd. 'de (2019) Türkiye örnekleminin olumlu yönde ve düşük etkiye sahip olduğunu ifade etmiştir. Bu sonuçlar bizim bulgularımızla da örtüşmektedir.

Aile, hayatın her bölümünde çocuklar için önemli bir unsurdur. Bir ebeveynin eğitimi beraberinde ebeveyn mesleğini ve ebeveyn gelirini etkilemektedir. Ebeveyn eğitimi öğrencinin okuma becerisinde önemli bir unsurdur (Geske ve Ozola 2008; Le vd. 2019; Rojas-Torres vd. 2021; Son vd. 2020; Vázquez-Cano, De la Calle-Cabrera, vd. 2020). Bu çalışmanın sonuçlarına göre de ebeveynin eğitimi önemli bir değişken olarak belirlenmiştir.

Dong ve Hu (2019) çalışmalarında okuma becerisini etkileyen indislerden biri olarak öğrencilerin test diline (LMINS) ayrılan süreyi belirlemiş olmakla birlikte, bizim çalışmamızda fen bilimleri ve toplam olarak haftalık alınan ders süreleri ön plana çıkmıştır. Nadaf vd. (2021) çalışmalarında diğer ders sürelerini analizden çıkardıktan sonra toplam öğrenme süresinin (TMINS) öğrencinin bilişsel başarısına katkıda bulunan en önemli özellik olduğunu tespit etmişlerdir. Eğitim süresinin haftada 35 saatten fazla olması durumunda sonuçlar üzerinde daha az olumlu ve hatta olumsuz etkiye sahip olduğunu belirtmişlerdir. Chung vd. (2021) çalışmalarında Koreli öğrencilerde TMINS önemli bir değişken olurken, Finlandiyalı öğrencilerde öğrenme sürelerinde etkinlik alt-başlıklarının (boş zaman, kendi başına okuma süresi vb) öne çıktığını belirtmişlerdir. Depren ve Depren (2021) Türkiye'deki öğrencilerin fen konularına harcadıkları zamanın okuma becerisi üzerinde istatistiksel olarak anlamlı bir etkisi olduğunu belirtmiştir. Bu çalışmada da okuma becerisi ile fen okuryazarlığı arasında istatistiksel olarak pozitif ve yüksek bir ilişki tespit edilmiştir.

Aşılacağı eğitim sistemlerinin yüz yüze yapıldığı sınıflardan, günümüzün teknolojik gelişmeleriyle çevrimiçi kurslar, videolarla başlayan eğitimleri özellikle 2020 yılında başlayan COVID-19 süreci ile geldiğimiz noktada sanal sınıflarda eğitim yapmak zorunda kalınmış, BİT ve eğitim arasındaki ilişki zorunluluklardan dolayı da olsa çok önemli bir hale gelmiştir. Okuma kültürünün değiştiği bu dönemde öğrencilere ait okuma becerilerini artırabilmek amacıyla öğrencilerin internette karşılaşabilecekleri olumsuz durumlara karşı kendilerini koruyabilmeleri için üstbiliş stratejilerinin ve BİT okuryazarlığının üzerinde durulması gerekmektedir. Politika yapımcılar ve eğitimcilerin hızla gelişmekte olan teknolojiye ayak uydurabilen genç neslini nitelikli bir hale getirebilmesi için öğrencilerin internete, dijital cihazlara erişimini artırmalı ve kullanmasını öğretmelidir. Okullar arası farklılığı gidermenin yollarını aramalıdır. Değişim gösteren okuma alışkanlığına karşı durmak yerine değişime ayak uydurulmalı sesli kitap, podcast ve eğitici videolar gibi öğrencileri besleyecek teknolojik araçlar kullanılmalı ve cazip hale getirilmelidir. Ayrıca bu çalışmada yer verilmeyen cinsiyetler ve okul türleri dikkate alınarak yapılacak yeni çalışmalar okuma becerisinde ön plana çıkan değişkenlerin incelenmesi de politika yapımcılar ve eğitimciler için yararlı bulgular sağlayacağı düşünülmektedir.

References

- Abbasoğlu, B. (2020). Ortaokul öğrencilerinin akademik başarılarının eğitsel veri madenciliği yöntemleri ile tahmini. *Veri Bilimi*, 3(1), 1–10. <https://dergipark.org.tr/tr/download/article-file/1198399>
- Adeyokun, B. O., Adeyanju, E. O., & Onyenania, G. O. (2020). Influence of entertainment media, cognitive styles and demographic variables on students' reading habits in yaba college of technology secondary school, Yaba, Lagos. *Information Impact: Journal of Information and Knowledge Management*, 10(2). <https://doi.org/10.4314/ijjkm.v10i2.3>
- Ahmed, A. A. M., Deo, R. C., Ghahramani, A., Raj, N., Feng, Q., Yin, Z., & Yang, L. (2021). LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios. *Stochastic Environmental Research and Risk Assessment*, 35(9). <https://doi.org/10.1007/s00477-021-01969-3>
- Akkoyunlu, B., & Kurbanoglu, S. (2003). Öğretmen Adaylarının bilgi okuryazarlığı ve bilgisayar öz-yeterlik algıları üzerine bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 24(24), 1–10. <https://dergipark.org.tr/tr/pub/hunefd/issue/7812/102529>
- Aksu, G., & Güzeller, C. O. (2016). Classification of PISA 2012 mathematical literacy scores using decision-tree method: turkey sampling. *TED Eğitim ve Bilim*, 41(185). <https://doi.org/10.15390/EB.2016.4766>
- Anand, N., Sehgal, R., Anand, S., & Kaushik, A. (2021). Feature selection on educational data using Boruta algorithm. *International Journal of Computational Intelligence Studies*, 10(1). <https://doi.org/10.1504/IJCISTUDIES.2021.113826>
- Anastasiou, D., Sideridis, G. D., & Keller, C. E. (2020). The relationships of Socioeconomic factors and special education with reading outcomes across PISA countries. *Exceptionality*, 28(4), 279–293. <https://doi.org/10.1080/09362835.2018.1531759>
- Arıkan, S., Özer, F., Şeker, V., & Ertaş, G. (2020). The Importance of sample weights and plausible values in large-scale assessments. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 522–539. <https://doi.org/10.21031/epod.602765>
- Arslan, A. (2020). Ortaokul öğrencilerinin matematiksel bilişüstü farkındalıklarının çeşitli değişkenler açısından belirlenmesi. *Turkish Journal of Educational Studies*, 7(2), 150–169. <https://dergipark.org.tr/tr/download/article-file/1108983>
- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 117–194. <https://doi.org/10.1016/j.compedu.2017.05.007>

- Bana, A. (2020). Students' Perception of using the internet to develop reading habits. *Journal of English Teaching*, 6(1), 60–70. <https://doi.org/10.33541/jet.v6i1.46>
- Bezek Güre, Ö., Kayri, M., & Erdoğan, F. (2020). Analysis of factors effecting pısa 2015 mathematics literacy via educational data mining. *TED Eğitim ve Bilim*, 45(202), 393–415. <https://doi.org/10.15390/EB.2020.8477>
- Breiman, L. (2001). Random forests. *Machine Learning volume*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Büyükkıdık, S., Bakırarar, B., & Bulut, O. (2018). Comparing the performance of data mining methods in classifying successful students with scientific literacy in PISA 2015. *The 6th International Congress on Measurement and Evaluation in Education and Psychology*, 68–75. <https://doi.org/10.7939/R3KW5812Q>
- Çelik, K., & Yurdakul, A. (2020). Investigation of PISA 2015 reading ability achievement of turkish students in terms of student and school level variables. *International Journal of Assessment Tools in Education*, 7(1), 30–42. <https://doi.org/10.21449/ijate.589280>
- Chung, H., Park, S., Kim, J.-I., & Kim, A. (2021). Exploring Variables affecting adolescents' reading literacy and life satisfaction: PISA 2018 International comparison of Korea and Finland. *Journal of Curriculum and Evaluation*, 24(1). <https://doi.org/10.29221/jce.2021.24.1.123>
- Çoban, Ö. (2020). Relationships between students' socioeconomic status, parental support, students' hindering, teachers' hindering and students' literacy scores: PISA 2018. *World Journal of Education*, 10(4), 45–59. <https://doi.org/10.5430/wje.v10n4p45>
- Cobb, D., & Couch, D. (2021). Locating inclusion within the OECD's assessment of global competence: An inclusive future through PISA 2018? *Policy Futures in Education*. <https://doi.org/10.1177/14782103211006636>
- Çocuk Vakfı. (2006). *Türkiye'nin okuma alışkanlığı karnesi*. https://cocukvakfi.org.tr/wp-content/dosya/raporlar/13_okuma_aliskanligi_karnesi2006.pdf
- Consulting, K., & Trust, N. L. (2013). *Youth literacy and employability commission: the report of the all-party parliamentary literacy group*. https://cdn.literacytrust.org.uk/media/documents/2013_01_01_free_other_-_Youth_Literacy_and_Employability_Commission_final_report.pdf
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4). <https://doi.org/10.1016/j.dss.2010.06.003>
- Depren, S. K., & Depren, Ö. (2021). Cross-cultural comparisons of the factors influencing the high reading achievement in Turkey and China: Evidence from

- PISA 2018. *The Asia-Pacific Education Researcher*.
<https://doi.org/doi.org/10.1007/s40299-021-00584-8>
- Ding, H., & Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research*, 102. <https://doi.org/10.1016/j.ijer.2020.101566>
- Dong, X., & Hu, J. (2019). An Exploration of impact factors influencing students' reading literacy in singapore with machine learning approaches. *International Journal of English Linguistics*, 9(5). <https://doi.org/10.5539/ijel.v9n5p52>
- Emdadi, A., & Eslahchi, C. (2021). Auto-HMM-LMF: feature selection based method for prediction of drug response via autoencoder and hidden Markov model. *BMC Bioinformatics*, 22(1), 33. <https://doi.org/10.1186/s12859-021-03974-3>
- Engel, L. C., Rutkowski, D., & Thompson, G. (2019). Toward an international measure of global competence? A critical look at the PISA 2018 framework. *Globalisation, Societies and Education*, 17(2). <https://doi.org/10.1080/14767724.2019.1642183>
- Ertem, H. Y. (2021). Examination of Turkey's PISA 2018 reading literacy scores within student-level and school-level variables. *Participatory Educational Research*, 8(1), 248–264. <https://doi.org/10.17275/per.21.14.8.1>
- Fabunmi, F., & Folorunso, O. (2010). Poor reading culture: A barrier to students' patronage of libraries selected secondary school in ado local government area of Ekiti-State, Nigeria. *African Research Review*, 4(2). <https://doi.org/10.4314/afrev.v4i2.58357>
- Fırat, T., & Koyuncu, İ. (2020). Investigating reading literacy in pisa 2018 assessment. *International Electronic Journal of Elementary Education*, 13(2), 263–275. <https://doi.org/10.26822/iejee.2021.189>
- Flores-Mendoza, C., Ardila, R., Gallegos, M., & Reategui-Colareta, N. (2021). General intelligence and socioeconomic status as strong predictors of student performance in latin american schools: evidence from PISAitems. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.632289>
- Gajwani, J., & Chakraborty, P. (2021). Students' performance prediction using feature selection and supervised machine learning algorithms. Kacprzyk, J., & P. Warsaw (Ed.), *International Conference on Innovative Computing and Communications* (ss. 37–354). Springer Nature Singapore. https://doi.org/10.1007/978-981-15-5113-0_25
- Gamazo, A., & Martínez-Abad, F. (2020). An exploration of factors linked to academic performance in PISA 2018 through data mining techniques. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.575167>

- Geske, A., & Ozola, A. (2008). Factors influencing reading literacy at the primary school level. *Problems of Education in the 21st Century*, 6(71–77). <http://www.scientiasocialis.lt/pec/node/files/pdf/Geske.pdf>
- Gubbels, J., Swart, N. M., & Groen, M. A. (2020). Everything in moderation: ICT and reading performance of Dutch 15-year-olds. *Large-scale Assessments in Education*, 8(1), 1. <https://doi.org/10.1186/s40536-020-0079-0>
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182. <https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>
- Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. I. Guyon, S. Gunn, M. Nikravesh, & L. A. Zadeh (Ed.), *Feature Extraction* (ss. 1–25). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35488-8_1
- Han, H., Jahed Armaghani, D., Tarinejad, R., Zhou, J., & Tahir, M. M. (2020). Random Forest and Bayesian Network techniques for probabilistic prediction of flyrock induced by blasting in quarry sites. *Natural Resources Research*, 29(2), 655–667. <https://doi.org/10.1007/s11053-019-09611-4>
- Han, Z., He, Q., & von Davier, M. (2019). Predictive feature generation and selection using process data from PISA Interactive problem-solving items: An application of Random Forests. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02461>
- Henry, L. A. (2006). Searching for an answer: The critical role of new literacies while reading on the internet. *The Reading Teacher*, 59(7), 614–627. <https://doi.org/10.1598/RT.59.7.1>
- Hootsuite. (2021). *We are social 2021 Türkiye raporu*. <https://datareportal.com/reports/digital-2021-turkey>
- Hu, J., Dong, X., & Peng, Y. (2021). Discovery of the key contextual factors relevant to the reading performance of elementary school students from 61 countries/regions: insight from a machine learning-based approach. *Reading and Writing*. <https://doi.org/10.1007/s11145-021-10176-z>
- Hu, X., Gong, Y., Lai, C., & Leung, F. K. S. (2018). The relationship between ICT and student literacy in mathematics, reading, and science across 44 countries: A multilevel analysis. *Computers & Education*, 125, 1–13. <https://doi.org/10.1016/j.compedu.2018.05.021>
- Iman, A. N., & Ahmad, T. (2020, Şubat). Improving intrusion detection system by estimating parameters of Random Forest in Boruta. *2020 International Conference on Smart Technology and Applications (ICoSTA)*. <https://doi.org/10.1109/ICoSTA48221.2020.1570609975>
- Jalota, C., & Agrawal, R. (2019). Analysis of educational data mining using

- classification. *Machine Learning, Big Data, Cloud and Parallel Computing*, 243–247.
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8862214&tag=1>
- Jalota, C., & Agrawal, R. (2021). Feature Selection algorithms and student academic performance: A study. *International Conference on Innovative Computing and Communications*, 317–328. https://doi.org/10.1007/978-981-15-5113-0_23
- Kasap, Y., Doğan, N., & Koçak, C. (2021). PISA 2018’de Okuduğunu anlama başarısını yordayan değişkenlerin veri madenciliği ile belirlenmesi. *Manisa Celal Bayar University Journal of Social Sciences*, 19(4), 241–258. <https://doi.org/10.18026/cbayarsos.959609>
- Kaunang, F. J., & Rotikan, R. (2018, Ekim). Students’ academic performance prediction using data mining. *2018 Third International Conference on Informatics and Computing (ICIC)*. <https://doi.org/10.1109/IAC.2018.8780547>
- Keskin, H. K. (2014). Programme for international student assessment (PISA) reading competencies: A study of the factors in academic reading. *The Anthropologist*, 18(1), 171–181. <https://doi.org/10.1080/09720073.2014.11891533>
- Khorramdel, L., Pokropek, A., Joo, S.-H., Kirsch, I., ve Halderman, L. (2020). Examining gender DIF and gender differences in the PISA 2018 reading literacy scale: A partial invariance approach. *Psychological Test and Assessment Modeling*, 60(2), 179–231. https://www.researchgate.net/publication/342344680_Examining_gender_DIF_and_gender_differences_in_the_PISA_2018_reading_literacy_scale_A_partial_invariance_approach
- Koğar, E. Y. (2021). An Investigation of the mediating role of various variables in the effect of both gender and economic, social and cultural status on reading literacy. *International Journal of Progressive Education*, 17(1), 376–391. <https://doi.org/10.29329/ijpe.2021.329.24>
- Koyuncu, İ., & Gelbal, S. (2020). Comparison of data mining classification algorithms on educational data under different conditions. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 11(4), 325–345. <https://doi.org/10.21031/epod.696664>
- Kursa, M. B. (2020). *Package ‘Boruta’*. <https://cran.r-project.org/web/packages/Boruta/Boruta.pdf>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11). <https://doi.org/10.18637/jss.v036.i11>
- Lahouar, A., & Slama, J. B. H. (2015). Random forests model for one day ahead load forecasting. *IREC2015 The Sixth International Renewable Energy Congress*, 1–6. <https://doi.org/10.1109/IREC.2015.7110975>

- Le, T.-T.-H., Tran, T., Trinh, T.-P.-T., Nguyen, C.-T., Nguyen, T.-P.-T., Vuong, T.-T., Vu, T.-H., Bui, D.-Q., Vuong, H.-M., Hoang, P.-H., Nguyen, M.-H., Ho, M.-T., & Vuong, Q.-H. (2019). Reading habits, socioeconomic conditions, occupational aspiration and academic achievement in Vietnamese junior high school students. *Sustainability*, 11(18). <https://doi.org/10.3390/su11185113>
- Lee, Y.-H., & Wu, J.-Y. (2012). The effect of individual differences in the inner and outer states of ICT on engagement in online reading activities and PISA 2009 reading literacy: Exploring the relationship between the old and new reading literacy. *Learning and Individual Differences*, 22(3), 336–342. <https://doi.org/10.1016/j.lindif.2012.01.007>
- Lee, Y. (2018). A study on development of collaborative problem solving prediction system based on deep learning: focusing on ICT factors. *Journal of The Korean Association of Information Education*, 22(1), 151–158. <https://doi.org/10.14352/jkaie.2018.22.1.151>
- Lezhnina, O., & Kismihók, G. (2021). Combining statistical and machine learning methods to explore German students' attitudes towards ICT in PISA. *International Journal of Research & Method in Education*, 1–20. <https://doi.org/10.1080/1743727X.2021.1963226>
- Liaw, A., ve Wiener, M. (2018). *Package 'randomForest'*. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Mahajan, G., ve Saini, B. (2020). Educational data mining: A state-of-the-art survey on tools and techniques used in EDM. *International Journal of Computer Applications & Information Technology*, 12(1), 310–316. <https://ijcait.com/IJCAIT/121/IJCAIT1215GINIKA.pdf>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, 100875. <https://doi.org/10.1016/j.stueduc.2020.100875>
- Mostafa, T. (2021). *Do girls and boys engage with global and intercultural issues differently?* <https://doi.org/10.1787/9a52e7dd-en>
- Muñoz, I. A., Molina, E. C., Casas, E. E., ve Martín, E. L. (2018). ¿Cuánto oro hay entre la arena? Minería de datos con los resultados de España en PISA 2015. *Revista Española de Pedagogía*, 76(270), 225–246. <https://www.jstor.org/stable/26547069>
- Mushtaq, S., Soroya, S. H., & Mahmood, K. (2020). Reading habits of generation Z students in Pakistan: Is it time to re-examine school library services? *Information Development*, 026666692096564. <https://doi.org/10.1177/0266666920965642>
- Nadaf, A., Eliëns, S., & Miao, X. (2021). Interpretable-machine-learning evidence for

- importance and optimum of learning time. *Int. J. Inf. Educ. Technol., Online First*, 1–6. <https://doi.org/10.18178/IJIE>
- Nzeyimana, G., & Bazimaziki, G. (2020). Revisiting the reading culture and information dissemination: Conceptualisation of “a reading nation is an informed nation”. *International Journal of English Literature and Social Sciences*, 5(3), 590–598. <https://doi.org/10.22161/ijels.53.5>
- OECD. (2016). *PISA 2018 Draft analytical frameworks*. <https://www.oecd.org/pisa/pisaproducts/PISA-2018-draft-frameworks.pdf>
- OECD. (2018). *Preparing our youth for an inclusive and sustainable world*. <https://www.oecd.org/pisa/Handbook-PISA-2018-Global-Competence.pdf>
- OECD. (2019a). PISA 2018 Assessment and analytical framework. *OECD iLibrary*. OECD.
- OECD. (2019b). *PISA database*. OECD. <https://www.oecd.org/pisa/data/>
- OECD. (2019c). PISA 2018 results (Volume III). *PISA 2018 Results (Volume III) What School Life Means for Students' Lives* (Annex A1.). OECD Publishing. <https://doi.org/10.1787/acd78851-en>
- OECD. (2020). Scaling procedures and construct validation of context questionnaire data. *PISA 2018 Technical Report* (s. 42). OECD Publishing. https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018_Technical-Report-Chapter-16-Background-Questionnaires.pdf
- Oriogu, C. D., Subair, R. E., Oriogu-Ogbuiyi, D. C., & Ogbuiyi, S. U. (2017). Effect of reading habits on the academic performance of students: A case study of the students of afe babalola university, Ado-Ekiti, Ekiti State. *Teacher Education and Curriculum Studies*, 2(5), 74. <https://doi.org/10.11648/j.tecs.20170205.13>
- Özkan, U. B. (2020). Öğrencilerde eudaimonianın ve akademik başarının yordayıcısı olarak ekonomik, sosyal ve kültürel düzey. *Yaşadıkça Eğitim*, 34(2). <https://doi.org/10.33308/26674874.2020342208>
- Özkan, U. B. (2021). Interest in Environmental issues as a determinant of science literacy: A multinational review with artificial neural network analysis. *FIRE: Forum for International Research in Education*, 7(1), 115–131. <https://doi.org/10.32865/fire202171232>
- Park, J., & Ranasinghe, W. M. D. T. (2021). A study on exploring digital information service method through analysis of PISA 2018 reading literacy assessment framework. *Journal of the Korean Society for Library and Information Science*, 55(1), 135–159. <https://doi.org/10.4275/KSLIS.2021.55.1.135>
- Pejic, A., & Molcer, P. S. (2019). Predicting the outcome of a PISA problem solving

- task using strategic behavior data. *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 313–318. <https://doi.org/10.1109/CogInfoCom47531.2019.9089942>
- Pejic, A., & Stanic Molcer, P. (2018). Relationship mining in PISA CBA 2012 problem solving dataset using association rules. *2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, 000549–000554. <https://doi.org/10.1109/SACI.2018.8440942>
- Petko, D., Cantieni, A., & Prasse, D. (2017). Perceived quality of educational technology matters. *Journal of Educational Computing Research*, *54*(8), 1070–1091. <https://doi.org/10.1177/0735633116649373>
- Pont, B., & Werquin, P. (2001). How old are new skills? *OECD Observer*, *225*, 15–17. <https://www.oecd-ilibrary.org/docserver/observer-v2001-2-en.pdf?expires=1623701762&id=id&accname=guest&checksum=B8E4146276B70DF42D8B1CBB2B2A091C>
- Rojas-Torres, L., Ordóñez, G., & Calvo, K. (2021). Teacher and Student practices associated with performance in the PISA reading literacy evaluation. *Frontiers in Education*, *6*. <https://doi.org/10.3389/feduc.2021.658973>
- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3). <https://doi.org/10.1002/widm.1355>
- Sağlam, Z., Pekyürek, M. F., & Yilmaz, R. (2020). PISA 2018 araştırmasına etki eden duygusal faktörlerin veri madenciliği yöntemleri ile incelenmesi. *Bilgi ve İletişim Teknolojileri Dergisi*, *2*(2), 113–148. <https://dergipark.org.tr/pub/bited/issue/58421/749242>
- Salal, Y. K., Abdullaev, S. M., & Kumar, M. (2019). Educational data mining: Student Performance prediction in academic. *International Journal of Engineering and Advanced Technology*, *8*(4C), 54–59. https://www.researchgate.net/publication/332369964_Educational_Data_Mining_Student_Performance_Prediction_in_Academic
- Sikora, J., & Pokropek, A. (2006). *Gendered career expectations of students*. <https://doi.org/http://dx.doi.org/10.1787/5kghw6891gms-en>
- Sokkhey, P., Navy, S., Tong, L., & Takeo, O. (2020). Multi-models of educational data mining for predicting student performance in mathematics: a case study on high schools in Cambodia. *IEIE Transactions on Smart Processing & Computing*, *9*(3), 217–229. <https://doi.org/10.5573/IEIESPC.2020.9.3.217>
- Son, Y., Hyunjeong, P., & Park, M. (2020). Random forest analysis of factors influencing the students' reading literacy levels: using PISA 2018 Korea data. *Asian Journal of Education*, *21*(1–4), 191–215. <https://doi.org/10.15753/aje.2020.03.21.1.191>

- Srijamdee, K., & Pholphirul, P. (2020). Does ICT familiarity always help promote educational outcomes? Empirical evidence from PISA-Thailand. *Education and Information Technologies*, 25(4), 2933–2970. <https://doi.org/10.1007/s10639-019-10089-z>
- Suna, E. H., Tanberkan, H., & Özer, M. (2020). Changes in literacy of students in turkey by years and school types: Performance of students in PISA applications. *Journal of Measurement and Evaluation in Education and Psychology*, 11(1), 76–98. <https://doi.org/10.21031/epod.702191>
- Suna, E., Tanberkan, H., Taş, E., Eroğlu, E., & Ümare, A. (2019). *PISA 2018 Türkiye ön raporu*. http://www.meb.gov.tr/meb_iys_dosyalar/2019_12/03105347_PISA_2018_Turkiye_On_Raporu.pdf
- Tat, O., Koyuncu, İ., & Gelbal, S. (2019). The influence of using plausible values and survey weights on multiple regression and hierarchical linear model parameters. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 235–248. <https://doi.org/10.21031/epod.486999>
- Tedmem. (2019). *Küresel yeterlik*. <https://tedmem.org/mem-notlari/degerlendirme/kuresel-yeterlik>
- Torppa, M., Eklund, K., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Why do boys and girls perform differently on PISA Reading in Finland? The effects of reading fluency, achievement behaviour, leisure reading and homework activity. *Journal of Research in Reading*, 41(1), 122–139. <https://doi.org/10.1111/1467-9817.12103>
- UNESCO. (2013). *Second global report on adult learning and education (GRALE 2)*. UNESCO Institute for Lifelong Learning. <http://www.unesco.org/new/fileadmin/MULTIMEDIA/FIELD/Santiago/pdf/GRALE2-Literacy-Chapter.pdf>
- Unpingco, J. (2019). Machine learning. *Python for Probability, Statistics, and Machine Learning* (2. baskı, s. 384). Springer International Publishing. <https://doi.org/10.1007/978-3-030-18545-9>
- Vázquez-Cano, E., De la Calle-Cabrera, A. M., Hervás-Gómez, C., & López-Meneses, E. (2020). Socio-family context and its influence on students' PISA reading performance scores: evidence from three countries in three continents. *Educational Sciences: Theory & Practice*, 20(2). <https://doi.org/10.12738/jestp.2020.2.004>
- Vázquez-Cano, E., Gómez-Galán, J., Infante-Moro, A., & López-Meneses, E. (2020). Incidence of a non-sustainability use of technology on students' reading performance in PISA. *Sustainability*, 12(2), 749.

<https://doi.org/10.3390/su12020749>

- Venkata, M. D., ve Lingamgunta, S. (2020). Breast cancer multi modality image analysis using phenotype features by SVM. *Journal of Science and Technology*, 5(1), 52–60. <http://jst.org.in/wp-content/uploads/2020/03/7.-Breast-Cancer-Multi-Modality-Image-Analysis-Using-Pheneotype-features-by-SVM.pdf>
- West, M., & Chew, H. E. (2014). *Reading in the mobile era: a study of mobile reading in developing countries*. R. Krau (Ed). UNESCO Institute for Lifelong Learning. https://unesdoc.unesco.org/ark:/48223/pf0000227436_eng
- Xiao, Y., Liu, Y., & Hu, J. (2019). Regression analysis of ICT impact factors on early adolescents' reading proficiency in five high-performing countries. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01646>
- Xu, X., Gu, H., Wang, Y., Wang, J., & Qin, P. (2019). Autoencoder based feature selection method for classification of anticancer drug response. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00233>
- Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212, 353–363. <https://doi.org/10.1016/j.snb.2015.02.025>
- Yilmaz, A., Fer, S., Kelecioglu, H., Doğan, N., Yazıcı, N., Özyalçın Oskay, Ö., Yetkin Özdemir, İ. E., ve Batı, K. (2020). *PISA ve Türkiye (2000 - 2018)*. <http://www.egitim.hacettepe.edu.tr/belge/pisaveturkiye.pdf>
- Zaffar, M., Hashmani, M. A., & Savita, K. S. (2017, Kasım). Performance analysis of feature selection algorithm for educational data mining. *2017 IEEE Conference on Big Data and Analytics (ICBDA)*. <https://doi.org/10.1109/ICBDAA.2017.8284099>

Ethical Declaration and Committee Approval

The study used the dataset provided by the OECD (2019b) to all researchers; thus, an ethical statement is not required.

The study utilized the data set made available to all researchers by OECD (2019b). Therefore, there is no need for an ethical declaration for the study.

Annex:**Table Annex 1.***Some literature studies on reading skills within the scope of PISA data*

Researchers	Year	Research Topics	Samples
Torppa et al. (2018)	PISA 2009, 2012	The effect of gender on reading skills	Finland
Xiao et al. (2019)	PISA 2015	The effect of ICT on reading skills	Singapore, New Zealand, Avusturalya, Finland and France
Vázquez-Cano et al. (2020)	PISA 2015	The impact of unsustainable technology use on students' reading skills	
Srijamdee and Pholphirul (2020)	PISA 2015	The impact of ICT familiarity on reading, science and mathematics literacy.	Thailand
Khorramdel et al. (2020)	PISA 2018	The effect of gender on reading skills	
Gamazo and Martínez-Abad (2020)	PISA 2018	Factors affecting student performance in reading, science and mathematics literacy	
Anastasiou et al. (2020)	PISA 2009 , 2010	The relationship between socioeconomic level and special education and reading, science and mathematics literacy	
Çoban (2020)	PISA 2018	Relationships between reading skills, socioeconomic status, parental support, student achievement and school level variables.	Turkiye
Özkan (2020)	PISA 2018	The impact of socioeconomic level on students' welfare levels and academic achievement in reading, science and mathematics literacy.	Turkiye

Suna et al. (2020)	PISA 2015, 2018	Change in reading, science and mathematics literacy according to years and school types	Turkiye
Son et al. (2020)	PISA 2018	Determination of variables that affect reading skills	Korea
Ding and Homer (2020)	PISA 2009	The importance of reading performance in explaining mathematics performance	China, Beijing
Gubbels vd (2020)	PISA 2015	The relationship between reading skills and ICT	Holland
Yurdakul and Çelik (2020)	PISA 2015	The relationship between students' and schools' reading skills	Turkiye
Vázquez-Cano et al. (2020)	PISA 2015	Reading skills of students from three countries with different socioeconomics were examined.	Canada, Finland and Singapore
Fırat and Koyuncu (2021)	PISA 2018	Determination of factors affecting reading skills	Turkiye
Flores-Mendoza et al. (2021)	PISA	The effect of general intelligence and socioeconomic level on reading, science and mathematics literacy	Latin American countries (Argentina, Brazil, Chile, Colombia and Peru)
Koğar (2021)	PISA 2009 , 2018	Examining the effects of gender and socioeconomic status on reading skills, with various variables acting as mediators.	Turkiye
Rojas-Torres et al. (2021)	PISA 2018	The evaluation of reading skills is done from the perspective of both teachers and students.	Costa Rica
Ertem (2021)	PISA 2018	Factors that affect reading skills are determined at both the student and school levels.	Turkiye

Depren and Depren (2021)	PISA 2018	A comparative examination of the factors affecting reading skills is conducted.	Turkiye, China
Park and Ranasinghe (2021)	PISA 2018	Examination of the relationship between reading skills and digital information services	
Chung et al. (2021)	PISA 2018	Examining the relationships between reading skills and life satisfaction	Korea, Finland

Table Annex 2.*Some studies in the literature on EDM*

Researchers	Research Topics	Datasets	Method
Aksu and Güzeller (2016)	The impact of students' interest in the course, attitude, motivation, perception, self-efficacy, anxiety, and study discipline on mathematical performance.	PISA 2012 Turkiye	Decision Trees (DT)
Asif et al. (2017)	Performance analysis	Undergraduate Students	DT and Clustering
Büyükkıdık et al. (2018)	Adequacy of science literacy	PISA2015 Turkiye	Multilayer perceptron (MLP), Logistic Regression (LR) and Support Vector Machines (SVM).
Muñoz et al. (2018)	Determination of factors associated with science, mathematics and reading performances	PISA2015 Spain	DT
Dong and Hu (2019)	Determining the characteristics that affect reading skills	PISA 2015 Singapore	SVM
Han et al. (2019)	Examining students' problem solving skills	PISA 2012	Random Forests (RF)
Martínez-Abad et al.(2020)	The effectiveness of the school on students	PISA2015 Spain	Hierarchical Modeling and DT
Son et al. (2020)	Determination of variables affecting reading skills	PISA 2018 Korea	RF

Koyuncu and Gelbal (2020)	Examining mathematics performance	PISA 2012 Turkiye	Naive Bayes (NB), K-nearest Neighborhood (K-NN), Artificial Neural Networks (ANN) and LR
Bezек Güre et al. (2020)	Factors affecting mathematical literacy	PISA 2015 Turkiye	RF and MLP Structural Equation Modeling (SEM), ANN, Sequential Minimal optimization, LR, K-NN, DT, RF and Deep Learning
Sokkhey et al. (2020)	Factors affecting student performance	Cambodian high school students	DT
Gamazo and Martínez-Abad (2020)	Factors affecting student performance in reading, science and mathematics literacy	PISA 2018	DT
Abbasoğlu (2020)	Prediction of academic success	Turkiye -Yalova Secondary School students	LR, SVM, Non-Linear SVM, RF, NB, K-NN, ANN
Sağlam et al. (2020)	Determining Emotional Factors	PISA 2018	Correlation and DT
Lezhnina and Kismihók (2021)	Exploring attitudes towards ICT	PISA 2015 and 2018 Germany	Multilevel Modeling and RF
Depren and Depren (2021)	Comparison of factors affecting reading skills	PISA 2018 Turkiye and China	Activity Region Finder (ARF)
Chung et al. (2021)	The relationship between reading skills and life satisfaction	PISA 2018 Korea and Finland	RF
Nadaf et al. (2021)	Determination of variables affecting student cognitive achievement Reading performance of primary school students from 61 countries/regions	PISA 2018	The SHapley Additive exPlanations (SHAP)
Hu et al. (2021)	Factors affecting student performance in reading, science and mathematics literacy	PIRLS 2016	SVM iterative feature elimination (SVM-RFE)

