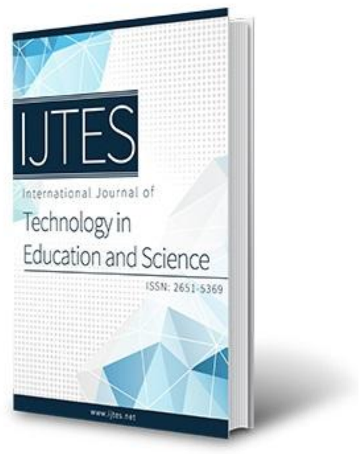

www.ijtes.net

Estimating Students' Adjustment Level in Distance Education Using Machine Learning and Resampling

Ayşe Alkan
Samsun Science and Art Center, Turkiye

To cite this article:

Alkan, A. (2025). Estimating students' adjustment level in distance education using machine learning and resampling. *International Journal of Technology in Education and Science (IJTES), 9*(1), 105-127. https://doi.org/10.46328/ijtes.570

IJTES
International Journal of Technology in Education and Science
2025, Vol. 9, No. 1, 105-127 https://doi.org/10.46328/ijtes.570

Estimating Students' Adjustment Level in Distance Education Using Machine Learning and Resampling

Ayşe Alkan

Article Info

Article History

Received:
25 April 2024

Accepted:
9 November 2024

Keywords

Online learning
Distance education
Machine learning
Resampling

Abstract

COVID-19's impact catalyzed the integration of distance education into our lives, shaping a crucial facet of modern learning. Initially, students, educators, and administrators encountered diverse challenges in navigating this paradigm shift. Swiftly addressing these hurdles promises enhanced efficacy in remote education. This research, blending experimental and descriptive methodologies, scrutinizes the "Students Adaptability Level in Online Education" dataset. It aims to assess students' adaptability in distance learning using five distinct machine learning techniques and identify pivotal factors influencing adaptation. Multiple classification endeavors aim to bolster predictive accuracy. Leveraging 14 resampling approaches, 70 classifications per algorithm—both with and without sampling—were conducted, each meticulously evaluated using four performance metrics. The Random Forest model, coupled with KMeansSMOTE oversampling, yielded a notable 93% accuracy, showcasing heightened classifier efficacy through resampling. Noteworthy correlations emerged, indicating that lesson durations of 1-3 hours, reliable internet connectivity, and financial assistance to families correlate with enhanced student adaptation. This study underscores the potential of resampling techniques in refining classification accuracy and underscores actionable strategies for optimizing distance education's effectiveness.

Introduction

There have been great changes in people's lives due to the earthquake that started with the COVID-19 pandemic and is now known as the Disaster of the Century, affecting eleven provinces in Turkey. After the health sector, one of the areas most affected by the changes is education (Yamamoto & Altun, 2020). The use of information technologies in education has not only enabled the use of technological tools and equipment in educational environments, but also brought new alternatives by moving away from the traditional education understanding, and has ensured the disappearance of time and space boundaries.

The use of computers in educational environments has enabled the use of new communication channels in a wide range of distance education. Distance education, which has started to take place more in our lives, is not a new education model. The emergence of distance education applications goes back about 300 years (Clark, 2020).

Today, the development of internet and mobile technologies has also provided an increase in distance education applications (Chang et al., 2018). Although various interpretations exist in academic discourse, distance education can be broadly understood as a pedagogical approach leveraging information technologies to facilitate learning across diverse temporal and spatial contexts (Valentine, 2002). Alternatively, it is characterized as an educational modality wherein learners and instructors operate within distinct environments, often employing tailored materials and resources to deliver course content (Usun, 2006).

It is necessary to systematically plan and make the necessary instructional design in order to ensure that the education of the students is not interrupted and the distance education process can be successful (Agormedah et al., 2020). Otherwise, some negativities may be experienced due to the adoption of traditional methods in educational institutions (Dhanarajan, 2001). In order to get the highest efficiency in the distance education process, it is necessary to eliminate the negativities. When the literature on the causes of the negativities experienced is examined; As a result of the research conducted by Hebebci, Bertiz, & Alan (2020), there are issues such as limited interaction, infrastructure and hardware problems among the prominent negative opinions about distance education (Hebebci et al., 2020). As a result of the study of Durak & Çankaya (2020), it was seen that the worries of those who took live lessons from university students who had anxiety before distance education and who used an integrated system completely disappeared (Durak et al., 2020). Gonzalez et al. (2020) also stated that distance education positively affects student achievement thanks to new learning methodologies (Gonzalez et al., 2020). Özdoğan & Berkant (2020) conducted research during the pandemic period, gathering insights from 137 stakeholders through semi-structured interviews. Stakeholders predominantly suggested solutions focusing on assessing and monitoring the process, ensuring equitable opportunities, enhancing engagement and interaction in lessons, bolstering infrastructure, utilizing in-house teachers for instruction, and streamlining lesson schedules (Özdoğan & Berkant, 2020). Meanwhile, Gençoğlu and Çiftçi (2020) explored the challenges faced by the education system amidst the global COVID-19 crisis, examining both worldwide and Turkish contexts, along with the measures and solutions devised. Their findings highlight key areas of concern across nations, primarily revolving around issues such as access to distance learning, student performance evaluation, formulation of compensatory education strategies, provision of psychosocial support, and addressing the needs of disadvantaged students (Gençoğlu & Çiftçi, 2020). Similarly, Chatterjee & Chakraborty (2020) surveyed students' perspectives on online education during the COVID-19 era, utilizing a 20-statement questionnaire. Notably, students expressed concerns about the stress induced by online learning and its impacts on their health and social lives (Chatterjee & Chakraborty, 2021).

Precautions to be taken against negative situations that may occur before starting the distance education process are an important element that will ensure the successful management of this process. The knowledge and experience of field experts make a great contribution to achieving the highest level of efficiency from distance education. However, the right decisions may not always be made by the experts. For this reason, it is important for experts to develop different tools to increase the efficiency of distance education in terms of increasing the quality of education. A large amount of data is collected from different stakeholders in the field of education. Connections between these data piles need to be established. However, advanced support mechanisms are needed to successfully establish connections between these data piles (Palaniappan & Awang, 2008). Rapid developments in information technologies create new opportunities for the provision of these support mechanisms. Machine learning, which has been widely used in the field of education, offers opportunities to provide fast solutions. With machine learning, computers can learn from existing data and predict new situations.

In this proposed study:

- The 14 attributes in the dataset were first converted to a numeric value.
- Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RO), Naive Bayes (NB) models, which were found to be effective in the field of education in the literature review, were used. As a result of the classification, the algorithms were compared in terms of accuracy, precision, recall, f1-score metrics.
- Due to the unbalanced distribution of the data in the data set, different from the studies in the literature, two different resampling techniques were applied with each classifier as undersampling and oversampling.
- In the data set, resampling methods were applied to all the classifiers one by one, and performance measurements were made and compared.
- Effective characteristics of students' adaptation level were determined.

Related Works

In the literature, there are studies in the field of education using different machine learning techniques. In these studies, Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), Naive Bayes (NB), Decision Trees (DT), Gradient Boosting (GB) and Artificial Neural Network (ANN) models were used. Ibrahim & Rusli (2007) endeavored to forecast university students' graduation grades using Decision Trees (DT), Artificial Neural Networks (ANN), and Linear Regression Analysis (LRA) techniques. They reported an 80% accuracy rate for their methodologies.

Karabatak (2008) employed the Apriori algorithm and association analysis to predict final grades of university students on the Moodle platform, achieving an average accuracy of 95.5% with association analysis. Dekker, Pechenizkiy & Vleeshouwers (2009) utilized the J48 algorithm, a type of DT method, to forecast students' attendance status with an 80% success rate (Dekker et al., 2009). Delen (2010) employed DT, Support Vector Machines (SVM), ANN, and LR models to anticipate and elucidate reasons behind university freshmen attrition, all achieving approximately 80% accuracy. Mishra, Kumar & Gupta (2014) utilized J48 and Random Tree (RT) algorithms via WEKA data mining software, achieving 88.37% accuracy with J48 and 94.41% with RT. Sara, Halland, Igel and Alstrup (2015) aimed to predict high school students' dropout status, with Random Forest (RF) algorithm achieving the highest accuracy at 93.47%, followed by SVM, CART, and NB algorithms (Sara et al., 2015).

Sivakumar, Venkataraman & Selvaraj (2016) improved upon DT in their research, achieving a high estimation accuracy of 97.50% for predicting university students' continuation in education (Sivakumar et al., 2016). Schatzel, Callahan, Scott and Davis (2011) utilized Ward's method, a clustering algorithm, to explore the potential for university dropouts to return to education, offering suggestions based on clusters. Şen, Uçar & Delen (2012) employed ANN, SVM, LR, and DT methods, highlighting the significance of past placement test scores in predicting secondary school placement exam success two years later.

Djulovic and Li (2013) developed attendance prediction models for university freshmen using DT, NB, ANN, and Rule Induction algorithms, with Rule Induction model achieving the highest overall accuracy at 86.27%. Iam-On and Boongoen (2017) utilized KNN algorithm in cluster analysis to identify university students' dropout tendencies, noting high attendance among students with strong academic backgrounds. Chung & Lee (2019) employed the RO algorithm to estimate high school students' dropout status, reporting 95% accuracy with RF algorithm. Iatrellis, Savvasi, Fitsilis and Gerogiannis (2021) proposed a two-stage machine learning approach integrating unsupervised and supervised learning techniques for highly accurate predictions of higher education program outcomes. Badal & Sungkur (2022) reported RF classifier's superior performance with 85% and 83% accuracy respectively in grade and attendance prediction, analyzing student performance and online learning platform features.

Çakıt and Dağdeviren (2022) found Extreme Gradient Boosting (XGBoost) algorithm to have higher accuracy in estimating student placement percentage based on university's academic reputation, city facilities, and university amenities. Guleria & Sood (2022) proposed a career counseling framework for students, achieving 91.2% Recall and 90.7% F-Measure scores with NB for predictions, outperforming LR, DT, SVM, KNN, and Ensemble models. Chen & Zhai (2023) explored diverse application scenarios of machine learning methods in their study.

Hadj Kacem, Alshehri and Qaid (2022) assert that educators require an expert in quality and education to validate the outcomes of their courses. This paper introduces a machine learning methodology for assessing Course Learning Outcomes (CLOs). The primary objective of this study is to create a model capable of evaluating the quality of a CLO. The paper introduces a novel approach named CLOCML (Course Learning Outcome Classification using Machine Learning) to construct predictive models for CLO paraphrasing. A newly compiled dataset named CLOC (Course Learning Outcomes Classes) was gathered for this purpose and subsequently subjected to a preprocessing phase. The performance of four models in predicting CLO classification was compared, including Support Vector Machine (SVM), Random Forest, Naive Bayes, and XGBoost. SVM emerged as the most effective classification model, achieving an accuracy rate of 83% in detecting the CLO class.

Cardona et al.(2023) present a comprehensive examination of the scholarly literature, offering a systematic review focused on forecasting student persistence in higher education via machine learning algorithms. The investigation centers on metrics like dropout risk, attrition risk, and completion risk. The review contributes a scholarly viewpoint concerning the anticipation of student retention using machine learning, elucidating various pivotal discoveries. These include discerning the variables employed in prior research endeavors and elucidating the methodologies employed for predictive analysis.

Wang's (2023) research endeavors to foster the robust and sustainable progression of music education in China. Consequently, this paper employs the aforementioned algorithms within the realm of classical music education. This encompasses tasks such as the identification of classical instruments, feature extraction, music recognition, and the assessment of classical music education quality. The effectiveness of the music quality evaluation system is determined by gauging the correlation between output results and subjective evaluations. A higher correlation signifies a superior music quality evaluation method. Empirical experiments validate that DTW score alignment and end-to-end approaches excel in extracting classical music features and exhibit enhanced accuracy in identifying classical instruments.

Sanusi et al. (2023) conducted a systematic examination of the state of research on the integration of Machine Learning (ML) into K-12 teaching and learning. Their study delves into the current thematic focus and identifies gaps that warrant attention in future scholarly investigations. The research findings highlight several key observations: (a) a demand for additional ML resources in kindergarten to middle school and informal learning environments, (b) a need for expanded exploration into the integration of ML across diverse subject domains beyond computing, (c) an emphasis on pedagogical development over teacher professional development programs in existing studies, and (d) a call for increased scrutiny of the societal and ethical implications of ML in forthcoming research endeavors.

Sperling et al. (2022) delve into the exploration of the underlying reasons and applications through which machine learning and artificial intelligence (AI) are making inroads into educational settings. Their article details ethnographic fieldwork conducted in Sweden, where 22 teachers and over 250 primary education students experimented with a machine learning teaching aid in mathematics known as the "AI Engine." Employing an Actor-Network Theory framework, the analysis hones in on the interactions within the network of diverse actors connected through the AI Engine, conceptualized as an "obligatory passage point." The findings shed light on how actions and narratives unfold within the intricate ecosystem of human actors, compensating for unforeseen and undesirable algorithmic decisions made by the AI Engine. Meng and Ma (2023) introduced a methodologically sound approach to identify "True Test Cheaters" in the dataset, showcased the efficacy of employing machine learning (ML) techniques to pinpoint anomalous statistical patterns in examination data, and formulated an analytical framework for the assessment and real-time implementation of ML-based test data forensics. The study assesses classification accuracy and false negative/positive outcomes across various supervised ML methodologies.

Basnet, Johnson and Doleck (2022) offer an extensive exploration into the comparative predictive capabilities of deep learning and machine learning in their paper. The focus is on leveraging educational big data to forecast dropout rates in MOOCs. Their findings reveal that machine learning classifiers exhibit predictive performance comparable to deep learning classifiers. This study contributes to the advancement of our comprehension regarding the utilization of deep learning and machine learning in enhancing models for predicting dropout rates. Dolawattha, Premadasa and Jayaweera (2022) aim to assess the viability of the envisioned mobile learning framework for higher education. Their study introduces an innovative approach, employing a machine learning-based ensemble method with severity indexes to gauge the sustainability of the proposed mobile learning system. The outcomes demonstrate that the suggested system has successfully attained economic and pedagogical sustainability. Notably, the study's distinctive contribution lies in its emphasis on a novel machine learning methodology for evaluating the sustainability of the proposed mobile learning framework.

Zhou and Jiao (2023) investigated the utilization of the stacking ensemble machine learning algorithm for analyzing test-takers' item responses, response times, and augmented data to identify cheating behaviors. The study conducted a comparative analysis of the stacking method against two other ensemble techniques (bagging and boosting) and six individual non-ensemble machine learning algorithms. The research addressed challenges related to class imbalance and input features. Results from the study revealed that stacking, coupled with resampling and feature sets that included augmented summary data, generally exhibited superior performance in detecting cheating compared to its counterparts. The meta-model derived from stacking, employing discriminant analysis based on the top two base models—Gradient Boosting and Random Forest—demonstrated superior performance when compared to other machine learning algorithms investigated in the study.

Abdelhafez and Elmannai (2022) embarked on a research endeavor aimed at early-stage prediction of student failures in specific courses utilizing standards-based grading. Employing various machine learning techniques such as SVM, multilayer perceptron, NB, and DT, the study demonstrated the efficacy of these algorithms in accurately predicting student failures post the third week and prior to the dropout week. This research not only enriches our understanding of student performance across diverse courses but also equips faculty members with valuable insights to assist at-risk students. By identifying and focusing on these students early on, faculty members can provide the necessary support to enhance their performance and prevent academic failure.

Method

The aim of this study is to predict the adaptation level of students to distance education by using different classifiers and to determine the most effective features at the level of adaptation. The research includes classification studies for students' adaptation level to distance education by using the "Students Adaptability Level in Online Education" dataset and different machine learning algorithms. Experimental studies and necessary analyzes were made on the dataset. For classification processes, LR, SVM, KNN, RF, and NB methods were used and their performances were evaluated. One of the main goals in classification studies is to increase prediction success. For this reason, five different methods and fourteen different resampling methods for each method were applied in the study. In total, the success of seventy separate classification processes were examined and the effective features in the adaptation process were determined.

Dataset

In this study, which aims to evaluate the adaptation level of students to distance education, the "Students Adaptability Level in Online Education" dataset was used“. The publicly shared “Students Adaptability Level in Online Education” dataset was accessed from the Kaggle website, which helps users find and publish the datasets (Web1). The dataset created with the questionnaire contains socio-demographic information of school, college and university students. The dataset consists of 1205 samples. There are 14 attributes (Suzan vd., 2021). The compliance level class value has three values as low, medium and high. In the dataset, there are 480 low, 625 medium, 100 high level of compliance data. Details of the features found in the dataset are given in Table 1. The statistical information of the features in the dataset is also given in Table 2.

Table 1. Features of the Dataset

Feature Name	Description
Gender	[Girl, Boy]
Age	[1-5,6-10,11-15,16-20,21-25,26-30, 30+]
Education Level	[School, College, University]
Institution Type	[Non Government, Government]
IT Student	[No, Yes]
Location in Town	[No, Yes]
Load-shedding	[Low, High]
Financial Condition	[Poor, Mid, Rich]
Internet Type	[2G, 3G, 4G]
Network Type	[Mobile Data, Wifi]
Class Duration	[0,1-3,3-6]
Self LMS	[No, Yes]
Device	[Tab, Mobile, Computer]
Adaptivity Level	[Low, Moderate, High]

Table 2. Descriptive Statistics for Dataset Features

Feature	Mean	Standard Deviation	Min	Max
Gender	0.449793	0.497679	0.00	1.00
Age	2.122822	1.210359	0.00	5.00
Education Level	1.196680	0.722437	0.00	2.00
Institution Type	0.682988	0.465506	0.00	1.00
IT Student	0.252282	0.434503	0.00	1.00
Location in Town	0.775934	0.417139	0.00	1.00
Load-shedding	0.833195	0.372956	0.00	1.00
Financial Condition	0.341909	0.605302	0.00	2.00
Internet Type	0.423237	0.494277	0.00	1.00
Network Type	1.627386	0.515295	0.00	2.00
Class Duration	1.047303	0.548559	0.00	2.00
Self LMS	0.174274	0.379502	0.00	1.00
Device	0.890456	0.384003	0.00	2.00
Adaptivity Level	1.435685	0.642013	0.00	2.00

Correlation relationships of dataset features are given in Figure 1. When Figure 1 is examined, it is seen that Financial Condition, Institution Type and Class Duration have higher correlations with the target class. Since there is no feature that has a correlation of more than 0.5 with the target, there is no single dominant feature in estimation.

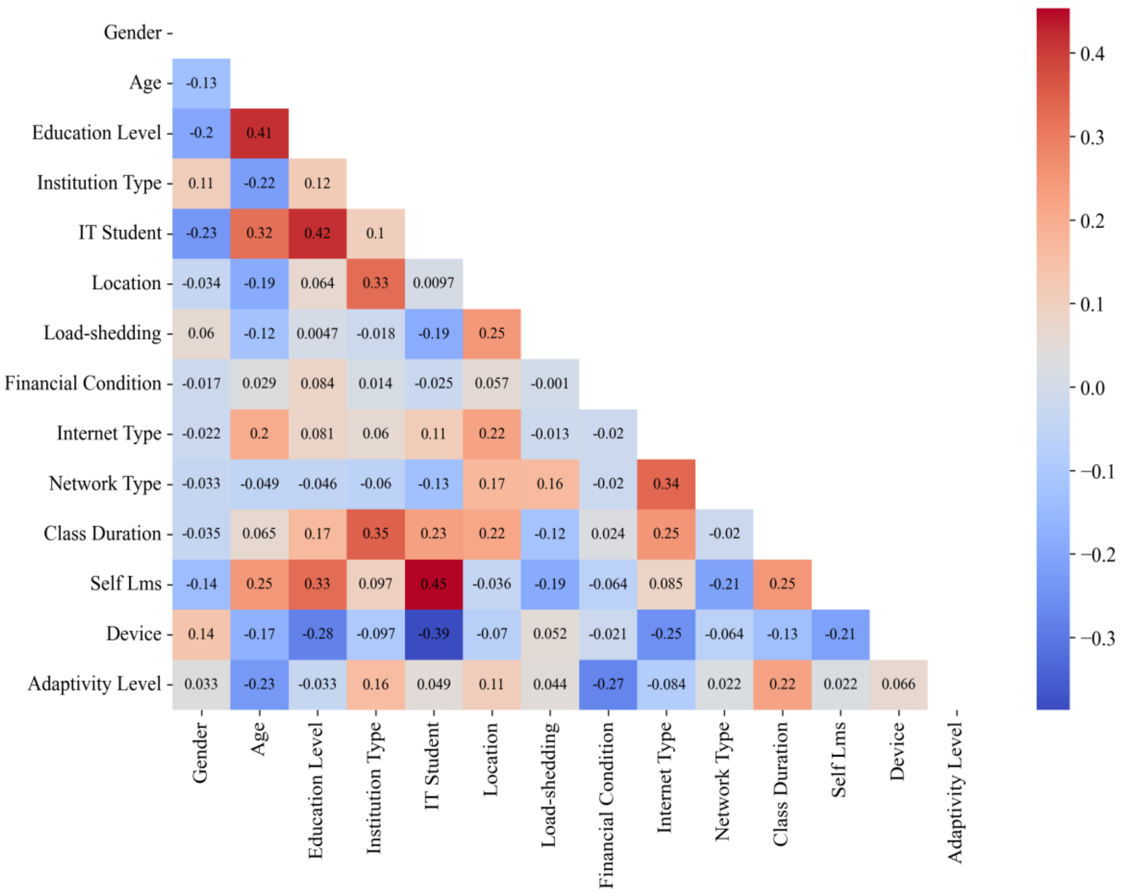

Figure 1. Correlation Matrix

Application Steps

By examining 14 features on the dataset, data distributions were analyzed and the features were converted into numerical values and made ready for experimental processing. LR, SVM, KNN, RF, NB classifiers were used in the data that was trained on the dataset. This is shown in Figure 2.

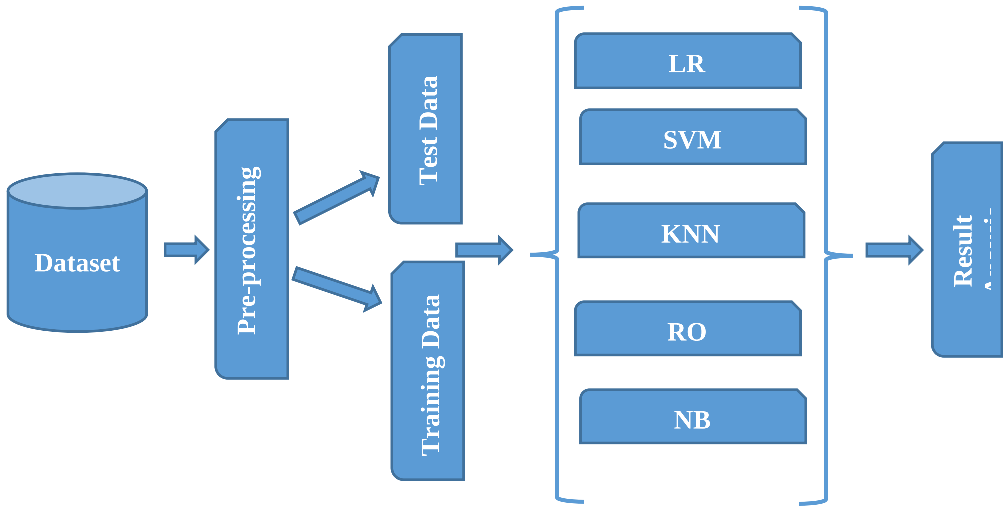

Figure 2. Sampleless Classification

The ability of the algorithms to predict the fit level feature was examined by comparing them in terms of accuracy, precision, recall, f1-score metrics. Since the data distribution in the dataset is not balanced, in order to increase the performance of the algorithms, the oversampling methods given in Figure 3 and the undersampling methods shown in Figure 4 were applied to all classifiers one by one and performance measurements were compared.

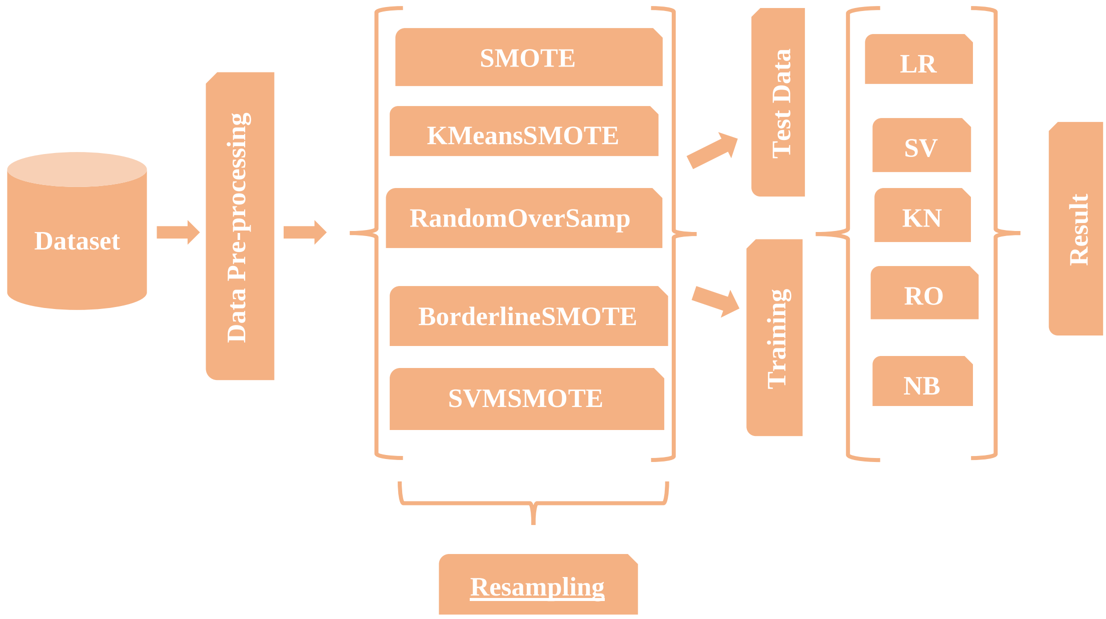

Figure 3. Classification by Oversampling

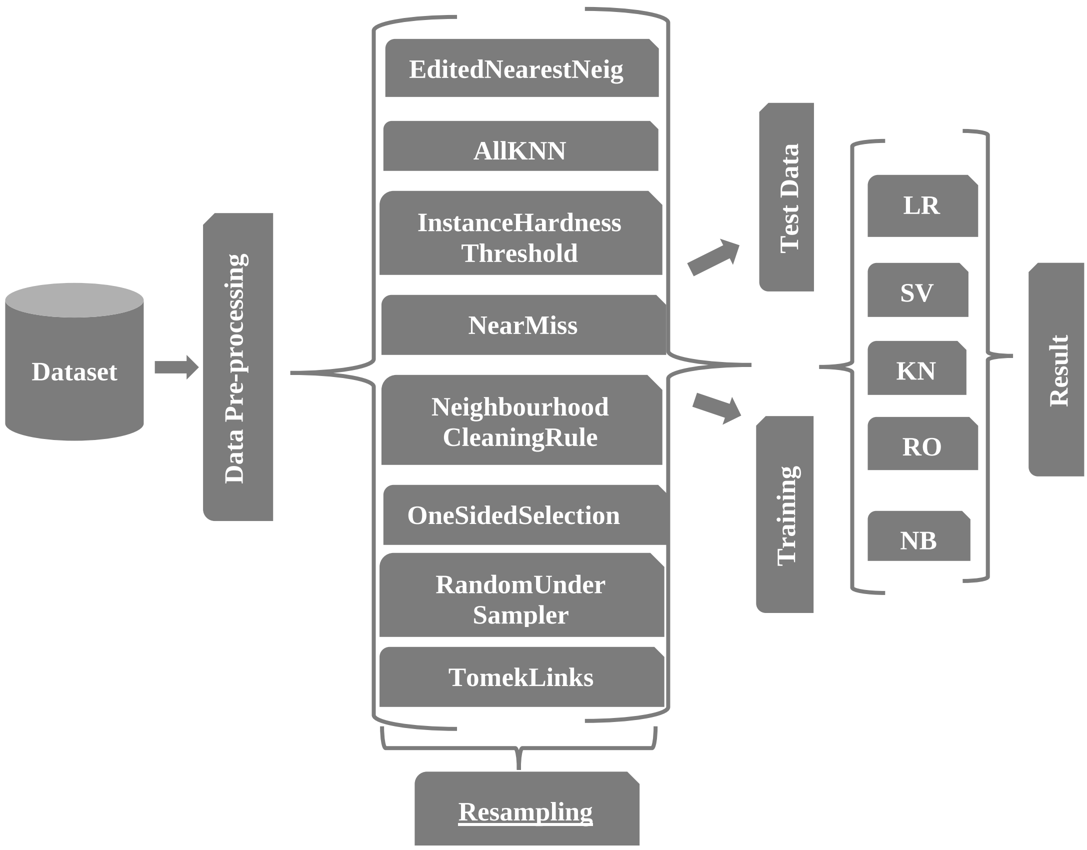

Figure 4. Classification by Undersampling

Classification Methods

Classification processes are the determination of which class the new observation belongs to as a result of the inferences made from the existing dataset. There are different methods that can be used for classification problems. In this study, LR, SVM, KNN, RF, NB classification methods, which are frequently used in classification studies in the literature, were used to determine the level of adaptation to distance education.

Support Vector Machine (SVM), Vapnik et al. It is a supervised machine learning technique based on statistical learning theory and used for classification and regression analysis (Vapnik vd., 1996). The Support Vector Machine (SVM) tries to maximize the margin between classes by classifying data points with a hyperplane that can be plotted in a space. When the SVM algorithm is used to distinguish between two classes, it finds the hyperplane that decides the best separation in the learning data. In addition, the SVM algorithm is particularly successful in high-dimensional datasets.

Logistic Regression algorithm (LR), Logistic regression analyzes the effect of an input variable on the output variable and expresses this effect as a probability value. Therefore, logistic regression analysis is used as a probability model to solve a classification problem. Logistic regression analysis is used to solve classification problems in many fields such as marketing, health, social sciences, economics and engineering.

K-Nearest Neighbor algorithm (KNN), It is a simple learning method that is particularly effective for small datasets. While the KNN algorithm can be used for classification or regression, it can give very successful results, especially when used for small datasets. However, as the dataset grows, the computation time increases and performance may degrade. Also, the KNN algorithm is prone to overfitting if not set correctly. The KNN algorithm is especially used in applications such as image, sound and natural language processing.

Random Forest algorithm (RF), RF based on the method developed by Ho (Ho, 1995)was later developed by Breiman (Breiman, 2001) and brought to the literature. RF is created by combining multiple decision tree models. The Random Forest algorithm is useful for reducing the risk of overfitting. It also delivers good results for many different data types and learns quickly. Random Forest algorithm is used to solve classification or regression problems in many fields, especially in image recognition, medical diagnosis, financial analysis, marketing and natural language processing.

Naive Bayes algorithm (NB), Naive Bayes analyzes the relationship between dependent and independent variables to derive a conditional probability (Aydoğan, 2008). The algorithm uses Bayes' theorem and assumes that all variables are independent given the value of the class variable(Dimitoglou vd., 2012). The algorithm is based on Bayes' theorem and probability concepts. The Naive Bayes algorithm calculates the probability that a data point belongs to a particular class, using the features and classes in the data set. Using Bayes' theorem, the algorithm calculates the conditional probability of the class of the data point, and then combines these probabilities to determine the most likely class.

Classification for Compliance Level in Distance Education

The features in the dataset, which consists of 14 attributes and 1205 samples, are numerical values. The class value is of numeric-categorical type as High (0), Low (1), Moderate (2). High, Low and Moderate distributions in the dataset are shown in Figure 5.

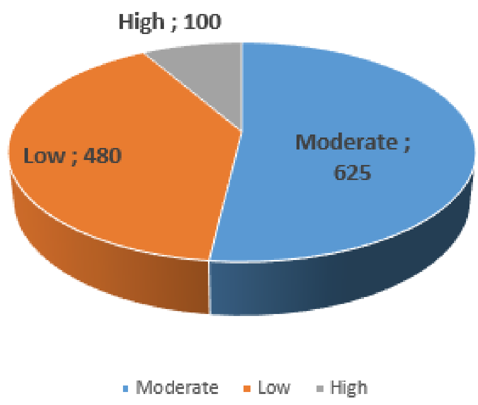

Figure 5. Distribution of Compliance Level in Distance Education in the Dataset

As shown in Figure 5, when the distribution of records of the level of adjustment in distance education is examined, it is seen that the number of those at the intermediate level is higher. It is seen that there is not a balanced distribution in the data distributions in the dataset. It is likely that the classification models used tend to be more inclined to this class, as the medium level enrollments have a rate of 51.86%.

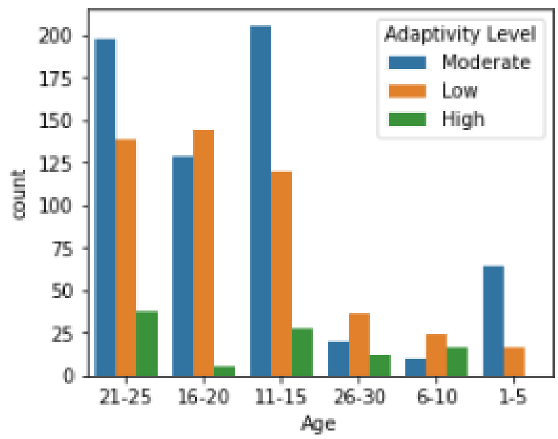

Figure 6**.** Distribution of Distance Education Adjustment Level by Age

Similarly, as shown in Figure 6, it is possible that the number of individuals between the ages of 21-25 and 11-15 years is close to each other, causing problems in the discrimination skills of the classification models. For this reason, resampling techniques were applied with each classifier. Resampling can be applied in two different ways: undersampling and oversampling. In this study, SMOTE (Chawla et al., 2002), KMeansSMOTE(Douzas et al., 2018), RandomOverSampler (Drummond et al., 2003), ADASYN(He et al., 2008), BorderlineSMOTE (Han et al., 2005) and SVMSMOTE(Nguyen et al., 2011); for undersampling, EditedNearestNeighbours (Wilson, 1972), AllKNN (Tomek, 1976a), InstanceHardnessThreshold (Smith et al., 2014), NearMiss (Mani & Zhang, 2003), NeighborhoodCleaningRule (Laurikkala, 2001), OneSidedSelection (Kubat et al., 1997) , RandomUnderSampler (Prusa et al., 2015) and TomekLinks (Tomek, 1976b) techniques were applied. The performance measurements were made by applying the resampling methods mentioned one by one to all the classifiers in the data set.

Performance Metrics

The values in the complexity matrix in Table 3 are used to evaluate the success of a classifier.

Table 3. Confusion Matrix

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Different success metrics are produced from the complexity matrix. The metrics used in this study are given in Table 4.

Table 4. Performance Metrics

Metrics	Mathematical expression of metrics
Accuracy	(TP + TN) / (TP + FP + FN + TN)
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1-Score	2 * Precision * Recall / (Precision + Recall)

The accuracy metric measures how accurate a model's predictions are. It represents the proportion of positive samples correctly classified by the precision metric, within the total positive predicted samples. The sensitivity metric gives the ratio of correctly classified positive samples to the total true positive samples, while the F1 score is a performance measure that combines precision and sensitivity metrics.

Results and Discussion

In this section, the experimental and descriptive results of the study are given.

Experimental Study Results

There is an unbalanced distribution in the dataset used in the study. In the dataset, SVM, LR, KNN, RF and NB classifiers and resampling techniques for each classifier were applied one by one and the performance measurements given in Table 4 were made. The parameters used in the study and their values are given in Table 5.

Table 5. Parameters used for the Classifiers and Their Values

Classifier	Parameters and values used
SVM	kernel='rbf', C=2
LR	max_iter=250
KNN	n_neighbors=13
RO	n_estimators =100
NB	Default

The results of each classification process applied without resampling and by doing it were measured with the metrics given in Table 4. The average values of the results were given by applying 5-fold cross validation in the classification processes. The results obtained are in Table 6 for LR, Table 7 for SVM, Table 8 for KNN, Table 9 for RF, Table 10 for NB, and Table 11 for the most successful results.

Table 6. Classification Results for Logistic Regression

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Without sample	-	0.71	0.7533	0.6566	0.6866
Oversampling	SMOTE	0.68	0.68	0.6766	0.6766
	Kmeans SMOTE	*0.79*	*0.7933*	*0.7833*	*0.7833*
	RandomOverSampler	0.68	0.69	0.6766	0.68
	ADYSN	0.65	0.65	0.65	0.65
	BorderLineSMOTE	0.61	0.6233	0.6133	0.6133
	SVMSMOTE	0.69	0.70	0.6966	0.6933
Undersampling	EditedNearestNeighbours	0.68	0.6933	0.68	0.68
	AllKNN	0.67	0.68	0.6633	0.67
	InstanceHardnessThreshold	0.65	0.66	0.65	0.65
	NearMiss	0.62	0.6333	0.6166	0.62
	NeighbourhoodCleaningRule	*0.68*	*0.6933*	*0.6833*	*0.6833*
	OneSidedSelection	0.68	0.69	0.6766	0.6766
	RandomUnderSampler	0.66	0.6666	0.66	0.6633
	TomekLinks	0.65	0.66	0.65	0.65

Table 7. Classification Results for Support Vector Machine

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Without sample	-	0.77	0.7966	0.6966	0.7266
Oversampling	SMOTE	0.78	0.7833	0.7766	0.7766
	Kmeans SMOTE	*0.81*	*0.8266*	*0.81*	*0.8033*

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
	RandomOverSampler	0.79	0.7933	0.7833	0.78
	ADYSN	0.75	0.7633	0.7533	0.7433
	BorderLineSMOTE	0.77	0.7733	0.7633	0.76
	SVMSMOTE	0.77	0.78	0.7666	0.77
Undersampling	EditedNearestNeighbours	0.78	0.7933	0.7766	0.7833
	AllKNN	0.77	0.7733	0.7633	0.77
	InstanceHardnessThreshold	*0.80*	*0.80*	*0.80*	*0.7966*
	NearMiss	0.79	0.79	0.7866	0.7833
	NeighbourhoodCleaningRule	0.76	0.7833	0.7633	0.77
	OneSidedSelection	0.79	0.79	0.79	0.79
	RandomUnderSampler	0.77	0.78	0.7733	0.7766
	TomekLinks	0.80	0.8033	0.7966	0.7933

Table 8. Classification Results for K Nearest Neighbor

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Without sample	-	0.79	0.76	0.6966	0.72
Oversampling	SMOTE	0.85	0.8433	0.8433	0.84
	Kmeans SMOTE	0.86	0.8666	0.8633	0.86
	RandomOverSampler	*0.88*	*0.8833*	*0.87*	*0.87*
	ADYSN	0.85	0.8666	0.8466	0.8433
	BorderLineSMOTE	0.84	0.8433	0.84	0.8433
	SVMSMOTE	0.85	0.85	0.8433	0.8433
Undersampling	EditedNearestNeighbours	0.86	0.8566	0.8533	0.8533
	AllKNN	0.85	0.85	0.8466	0.84
	InstanceHardnessThreshold	0.84	0.8366	0.8366	0.8333
	NearMiss	0.84	0.84	0.84	0.8433
	NeighbourhoodCleaningRule	0.85	0.8566	0.8533	0.85
	OneSidedSelection	*0.86*	*0.86*	*0.8566*	*0.8566*
	RandomUnderSampler	0.85	0.8466	0.8466	0.8433
	TomekLinks	0.85	0.8466	0.8433	0.84

Table 9. Classification Results for Random Forest

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Without sample	-	0.91	0.9033	0.8566	0.88

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Oversampling	SMOTE	0.91	0.9166	0.91	0.91
	Kmeans SMOTE	*0.93*	*0.9266*	*0.9233*	*0.9233*
	RandomOverSampler	0.92	0.9266	0.9233	0.9233
	ADYSN	0.89	0.8866	0.8866	0.8866
	BorderLineSMOTE	0.91	0.9133	0.9066	0.9066
	SVMSMOTE	0.91	0.9066	0.9066	0.9066
Undersampling	EditedNearestNeighbours	0.91	0.9066	0.9033	0.9033
	AllKNN	0.90	0.90	0.8966	0.8966
	InstanceHardnessThreshold	0.90	0.9033	0.90	0.90
	NearMiss	0.90	0.8966	0.8933	0.8933
	NeighbourhoodCleaningRule	*0.91*	*0.9166*	*0.9133*	*0.91*
	OneSidedSelection	0.91	0.9066	0.9033	0.9033
	RandomUnderSampler	0.91	0.91	0.9033	0.9066
	TomekLinks	0.91	0.91	0.91	0.9066

Table 10. Classification Results for Navie Bayes

Sampling	Technique Used	Accuracy	Precision	Recall	F1-Score
Without sample	-	0.70	0.6866	0.67	0.6766
Oversampling	SMOTE	0.63	0.6333	0.6333	0.6333
	Kmeans SMOTE	*0.74*	*0.7433*	*0.7366*	*0.7333*
	RandomOverSampler	0.62	0.6266	0.6233	0.62
	ADYSN	0.63	0.6333	0.6266	0.6266
	BorderLineSMOTE	0.64	0.6433	0.6366	0.6333
	SVMSMOTE	0.64	0.6366	0.64	0.64
Undersampling	EditedNearestNeighbours	0.60	0.6066	0.6066	0.60
	AllKNN	0.62	0.6166	0.6166	0.6133
	InstanceHardnessThreshold	0.62	0.62	0.6133	0.6166
	NearMiss	0.61	0.62	0.61	0.6066
	NeighbourhoodCleaningRule	*0.62*	*0.62*	*0.62*	*0.62*
	OneSidedSelection	0.61	0.6066	0.6066	0.6066
	RandomUnderSampler	0.60	0.6066	0.6066	0.6033
	TomekLinks	0.60	0.6033	0.6033	0.6033

When the measurements in the tables were evaluated, it was observed that some resampling techniques increased the success. Although undersampling and oversampling techniques varied in success, oversampling techniques yielded more successful results in this study. It has been seen that KMeansSMOTE and RandomOverSampler among oversampling techniques, and NeighborhoodCleaningRule, OneSidedSelection, and InstanceHardnessThreshold techniques among undersampling techniques provide higher success. All the classifiers used and the most successful ones of the sampling methods are summarized in Table 11.

Table 11. The Most Successful Results Achieved

Sampling	Technique Used	The best Accuracy	The best Precision	The best Recall	The best F1-Score
LR	Without sample	0.71	0.7533	0.6566	0.6866
	Oversampling	*0.79*	*0.7933*	*0.7833*	*0.7833*
	Undersampling	*0.68*	*0.6933*	*0.6833*	*0.6833*
SVM	Without sample	0.77	0.7966	0.6966	0.7266
	Oversampling	*0.82*	*0.84*	*0.82*	*0.81*
	Undersampling	*0.80*	*0.80*	*0.80*	*0.7966*
KNN	Without sample	0.79	0.76	0.6966	0.72
	Oversampling	*0.88*	*0.8833*	*0.87*	*0.87*
	Undersampling	*0.86*	*0.86*	*0.8566*	*0.8566*
RF	Without sample	0.91	0.9033	0.8566	0.88
	Oversampling	*0.93*	*0.9266*	*0.9233*	*0.9233*
	Undersampling	*0.91*	*0.9166*	*0.9133*	*0.91*
NB	Without sample	0.70	0.6866	0.67	0.6766
	Oversampling	*0.74*	*0.7433*	*0.7366*	*0.7333*
	Undersampling	*0.62*	*0.62*	*0.62*	*0.62*

When the results in Table 11 are examined, it is seen that high values are obtained with Random Forest resampling. The highest values of RF were obtained in terms of all parameters. The highest accuracy value obtained with RF was obtained with KmeansSMOTE oversampling technique. The best results in all classifications with RF were 93% for accuracy, 9266% for precision, 9233% for sensitivity and 9233% for F1 Score. After the RF classifier, the highest measurements were obtained with KNN.

The highest accuracy value obtained with this classifier was obtained with the RandomOverSampler oversampling technique. The best results for all classifications with KNN were 88% for accuracy, 8833% for precision, 87% for sensitivity and 87% for F1 Score. Within the scope of the study, the best accuracy values suggested from the models with the resampling method applied together with the LR, SVM, KNN, RF and NB classifiers are given in Table 12.

Table 12. Comparison of the Findings with the Studies in the Literature

Reference	Method(s)	The best accuracy (%)
(Suzan vd., 2021)	DT	87.56
	RF	89.63
	NB	70.95
	SVM	66.80
	KNN	76.348
	Artificial Neural Network (ANN)	82.99
	Literature Mean	79.04633
	Proposed LR	79
	Proposed SVM	82
	Proposed KNN	88
	Proposed RF	93
	Proposed NB	74
	Mean of Proposed Models	83.2

Studies on the dataset used in the study are limited in the literature. However, the resampling used in this study provided a higher success. In the previous study, the highest accuracy values were 89.63%, then 87.56% and 82.99%. The average success is 79.04633%. In this study; the highest accuracy values are 93%, then 88% and 82%. The average success is 83.2%. Within the scope of the study, the highest accuracy is followed by RF and then KNN. High success was achieved in RF, KmeansSMOTE oversampling method, and RandomOverSampler oversampling method in KNN.

Descriptive Study Results

When we examine the effect of 13 features in the dataset, other than the level of fit, on the prediction level of the classifiers, the effect levels of the features are shown in Figure 7.

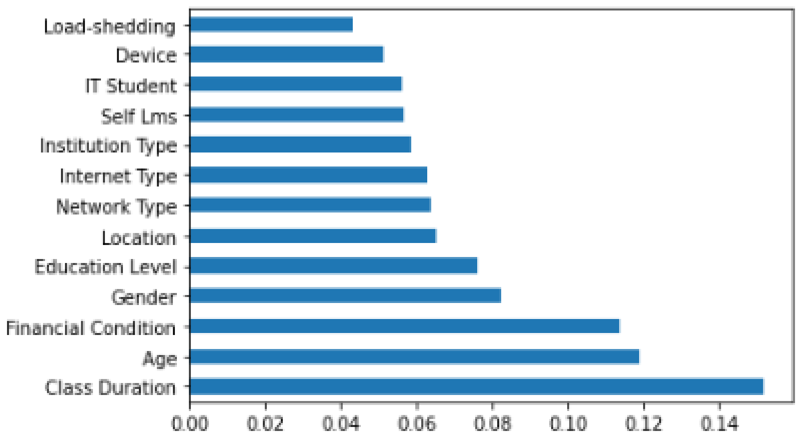

Figure 7. Distribution of the Most Important Features

When the distribution in Figure 7 is examined, it can be said that Class Duration, Age, Financial Condition, Gender and Education Level features are the first 5 features that are effective in determining the level of adaptation of students. The distributions of the 5 features related to the fit feature are shown in Figure 8. When the most effective features in determining the adaptation feature in Figure 8 are examined, it can be said that the duration of the course between 1-3 hours increases the success in the adaptation process to distance education. It is seen that students between the ages of 11-15, 16-20 and 21-25 are at a moderate level in the adaptation process. These levels can be increased with technology lessons to be given to students. Financial support to families can help establish a better connection. In addition, it can be said that the adaptation of students to distance education can be increased by providing internet connections to students studying in public schools.

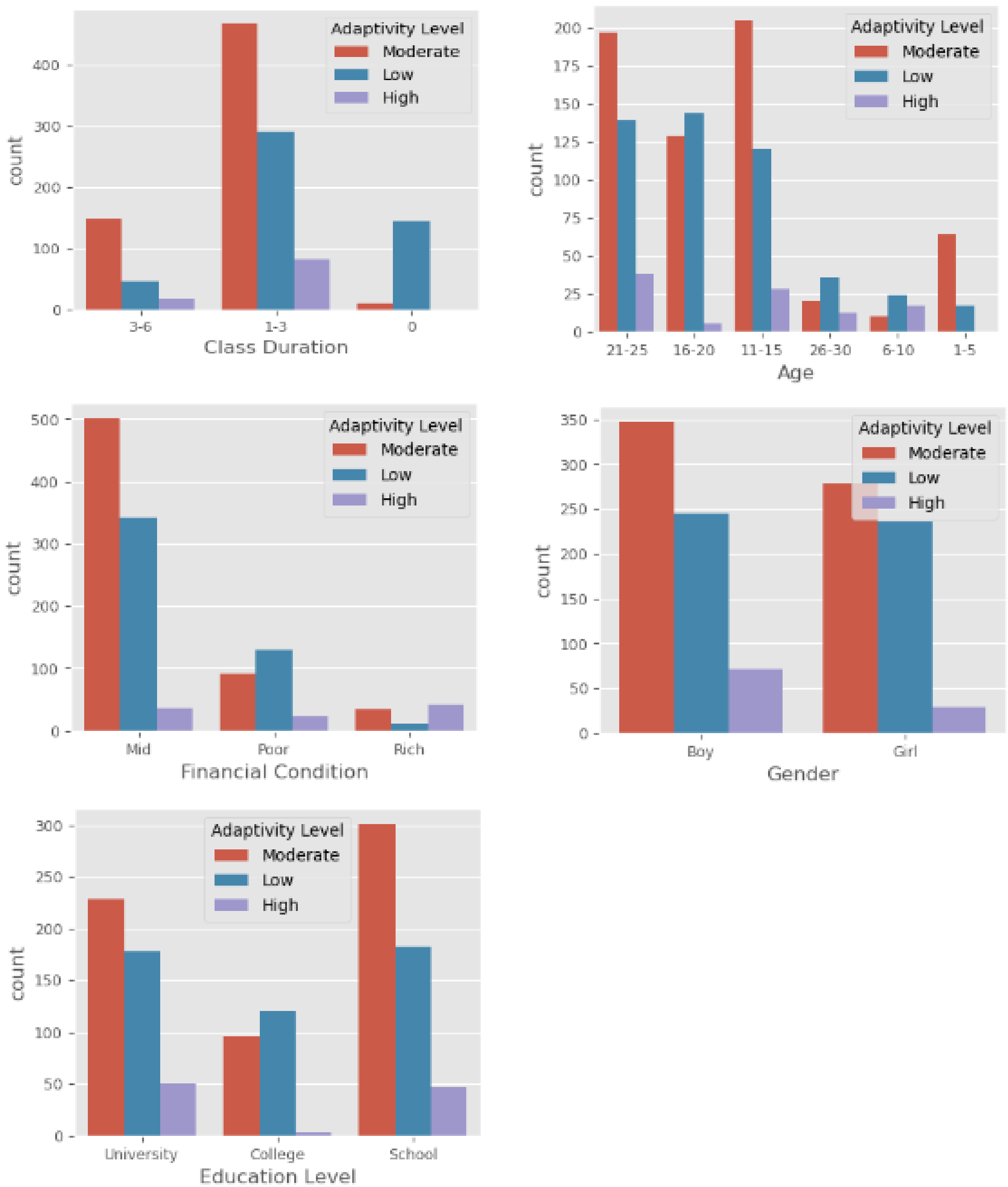

Figure 8. Distribution of the 5 Most Important Features According to the Adaptation Feature

Conclusions

It provides the opportunity to benefit from education in the most efficient way by revealing the critical patterns found in large data piles obtained in the field of education with machine learning. In this study, machine learning techniques were used to evaluate the adaptation level of individuals to distance education. In this study, classifications were made with five different machine learning methods in the "Students Adaptability Level in Online Education" dataset. Since the values in the "Students Adaptability Level in Online Education" dataset were not balanced, fourteen different resampling techniques were used with the five classifiers used. For LR, SVM, KNN, RF and NB machine learning models, the classification successes obtained without resampling and by performing fourteen separate resamplings were compared. It has been found that resampling methods with four metrics, namely accuracy, precision, sensitivity and F1 score, of a total of seventy separate classifications, increase the success of classification in general. In this study, 93% accuracy obtained with RF classifier using KmeansSMOTE oversampling technique and 88% accuracy higher than KNN classifier with RandomOverSampler oversampling method. It has been seen that resampling increases the success of the classifier in order to ensure the class balance in the dataset. KMeansSMOTE and RandomOverSampler, which are oversampling techniques used in this study, provided a higher success rate. The imbalances of the classes in the dataset is a factor that affects the success of the models. Balancing the existing dataset with appropriate sampling techniques allows the models to be more successful. In this study, the achievements of different machine learning models were compared and the results obtained by resampling techniques formed an opinion about the performance status of the models. In addition, the factors affecting the adaptation process of students in distance education were determined with machine learning techniques. It can be said that working on these elements and providing new opportunities to students will increase success. Researchers working in this field can contribute to the field by determining the factors affecting education and creating more adaptive educational environments.

Recommendations

Machine learning offers revolutionary innovations in education, making learning processes more efficient and effective. If we need to make recommendations to researchers in this area, it is important that they first investigate how to integrate data analysis and artificial intelligence techniques into education. They can maximize the potential of machine learning in education by focusing on applications such as predicting student performance, providing personalized learning experiences, and evaluating the effectiveness of educational materials. In addition, ensuring transparency in data collection and usage processes by paying attention to ethical and privacy issues also plays a critical role. In this way, the reliability and acceptability of machine learning-based educational technologies can be increased. Finally, working with experts from different fields such as educational sciences, computer science, and psychology by establishing interdisciplinary collaborations will contribute to the development of more comprehensive and effective solutions.

References

Abdelhafez, H. A., & Elmannai, H. (2022). Developing and comparing data mining algorithms that work best for predicting student performance. *International Journal of Information and Communication Technology Education (IJICTE)*, *18*(1), 1-14.

Agormedah, E. K., Henaku, E. A., Ayite, D. M. K., & Ansah, E. A. (2020). Online learning in higher education during COVID-19 pandemic: A case of Ghana. *Journal of Educational Technology and Online Learning*, *3*(3), 183-210.

Aydoğan, E. K. (2008). *Veri madenciliğinde sınıflandırma problemleri için evrimsel algoritma tabanlı yeni bir yaklaşım:rough-mep algoritması*. (Yayınlanmamış Doktora Tezi,149, Gazi Üniversity, Turkey.

Basnet, R. B., Johnson, C., & Doleck, T. (2022). Dropout prediction in Moocs using deep learning and machine learning. *Education and Information Technologies*, *27*(8), 11499-11513.

Badal, Y. T., & Sungkur, R. K. (2022). Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies*, 1-31.

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

Cardona, T., Cudney, E. A., Hoerl, R., & Snyder, J. (2023). Data mining and machine learning retention models in higher education. *Journal of College Student Retention: Research, Theory & Practice*, *25*(1), 51-75.

Chang, C.-Y., Lai, C.-L., & Hwang, G.-J. (2018). Trends and research issues of mobile learning studies in nursing education: A review of academic publications from 1971 to 2016. *Computers & Education*, *116*, 28-48.

Chatterjee, I., & Chakraborty, P. (2021). Use of information communication technology by medical educators amid COVID-19 pandemic and beyond. *Journal of Educational Technology Systems*, *49*(3), 310-324.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, 1-19.

Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, *96*, 346-353.

Clark, J. T. (2020). Distance education. İçinde *Clinical engineering handbook* (ss. 410-415). Elsevier.

Çakıt, E., & Dağdeviren, M. (2022). Predicting the percentage of student placement: A comparative study of machine learning algorithms. *Education and Information Technologies*, *27*(1), 997-1022.

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *International Working Group on Educational Data Mining*.

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498-506.

Dhanarajan, G. (2001). Distance education: Promise, performance and potential. *Open Learning: The Journal of Open, Distance and e-Learning*, *16*(1), 61-68.

Dimitoglou, G., Adams, J. A., & Jim, C. M. (2012). *Comparison of the C4.5 and a Naive Bayes Classifier for the Prediction of Lung Cancer Survivability* (arXiv:1206.1121). arXiv. https://doi.org/10.48550/arXiv.1206.1121

Djulovic, A., & Li, D. (2013). Towards freshman retention prediction: A comparative study. *International Journal of Information and Education Technology*, *3*(5), 494-500.

Dolawattha, D. M., Premadasa, H. S., & Jayaweera, P. M. (2022). Evaluating sustainability of mobile learning framework for higher education: a machine learning approach. *The International Journal of Information* *and Learning Technology*, *39*(3), 266-281.

Douzas, G., Bacao, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*, *465*, 1-20.

Drummond, C., Holte, R. C., & others. (2003). C4. 5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. *Workshop on learning from imbalanced datasets II*, *11*, 1-8.

Durak, G., ÇANKAYA, S., & İzmirli, S. (2020). COVID-19 pandemi döneminde Türkiye'deki üniversitelerin uzaktan eğitim sistemlerinin incelenmesi. *Necatibey Eğitim Fakültesi elektronik fen ve matematik eğitimi dergisi*, *14*(1), 787-809.

Gençoğlu, C., & Çiftçi, M. (2020). COVID-19 salgınında eğitim: Türkiye üzerinden bir analiz1. *Journal of History School (JOHS)*.

Gonzalez, T., De La Rubia, M. A., Hincz, K. P., Comas-Lopez, M., Subirats, L., Fort, S., & Sacha, G. M. (2020). Influence of COVID-19 confinement on students' performance in higher education. *PloS one*, *15*(10), e0239490.

Hadj Kacem, Y., Alshehri, S., & Qaid, T. (2022). Categorizing Well-Written Course Learning Outcomes Using Machine Learning. *Journal of Information Technology Education: Innovations in Practice*, *21*, 61-75.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23-26, 2005, Proceedings, Part I 1*, 878-887.

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 1322-1328.

Hebebci, M. T., Bertiz, Y., & Alan, S. (2020). Investigation of views of students and teachers on distance education practices during the Coronavirus (COVID-19) Pandemic. *International Journal of Technology in Education and Science*, *4*(4), 267-282.

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, *1*, 278-282.

Iam-On, N., & Boongoen, T. (2017). Generating descriptive model for student dropout: A review of clustering approach. *Human-centric Computing and Information Sciences*, *7*(1), 1-24.

Iatrellis, O., Savvas, I. K., Fitsilis, P., & Gerogiannis, V. C. (2021). A two-phase machine learning approach for predicting student outcomes. *Education and Information Technologies*, *26*, 69-88.

Ibrahim, Z., & Rusli, D. (2007). Predicting students' academic performance: Comparing artificial neural network, decision tree and linear regression. *21st Annual SAS Malaysia Forum, 5th September*.

Karabatak, M. (2008). *Özellik Seçimi, Sınıflama ve Öngörü Uygulamalarına Yönelik Birliktelik Kuralı Çıkarımı ve Yazılım Geliştirilmesi*. Fırat Üniversitesi.

Kubat, M., Matwin, S., & others. (1997). Addressing the curse of imbalanced training sets: One-sided selection. *Icml*, *97*(1), 179.

Laurikkala, J. (2001). Improving identification of difficult small classes by balancing class distribution. *Artificial Intelligence in Medicine: 8th Conference on Artificial Intelligence in Medicine in Europe, AIME 2001 Cascais, Portugal, July 1–4, 2001, Proceedings 8*, 63-66.

Mani, I., & Zhang, I. (2003). kNN approach to unbalanced data distributions: A case study involving information extraction. *Proceedings of workshop on learning from imbalanced datasets*, *126*, 1-7.

Meng, H., & Ma, Y. (2023). Machine Learning–Based Profiling in Test Cheating Detection. *Educational Measurement: Issues and Practice*, *42*(1), 59-75.

Mishra, T., Kumar, D., & Gupta, S. (2014). Mining students' data for prediction performance. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 255-262.

Nguyen, H. M., Cooper, E. W., & Kamei, K. (2011). Borderline over-sampling for imbalanced data classification. *International Journal of Knowledge Engineering and Soft Data Paradigms*, *3*(1), 4-21.

Özdoğan, A. Ç., & Berkant, H. G. (2020). Covid-19 pandemi dönemindeki uzaktan eğitime ilişkin paydaş görüşlerinin incelenmesi. *Milli Eğitim Dergisi*, *49*(1), 13-43.

Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. *2008 IEEE/ACS international conference on computer systems and applications*, 108-115.

Prusa, J., Khoshgoftaar, T. M., Dittman, D. J., & Napolitano, A. (2015). Using random undersampling to alleviate class imbalance on tweet sentiment data. *2015 IEEE international conference on information reuse and integration*, 197-202.

Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, *28*(5), 5967-5997.

Sara, N.-B., Halland, R., Igel, C., & Alstrup, S. (2015). High-school dropout prediction using machine learning: A Danish large-scale study. *ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence*, 319-324.

Schatzel, K., Callahan, T., Scott, C. J., & Davis, T. (2011). Reaching the non-traditional stopout population: A segmentation approach. *Journal of Marketing for Higher Education*, *21*(1), 47-60.

Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, *9*(4), 1-5.

Smith, M. R., Martinez, T., & Giraud-Carrier, C. (2014). An instance level analysis of data complexity. *Machine learning*, *95*, 225-256.

Sperling, K., Stenliden, L., Nissen, J., & Heintz, F. (2022). Still w (AI) ting for the automation of teaching: An exploration of machine learning in Swedish primary education using Actor-Network Theory. *European Journal of Education*, *57*(4), 584-600.

Suzan, M. H., Samrin, N. A., Biswas, A. A., & Pramanik, A. (2021). Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-7.

Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, *39*(10), 9468-9476.

Tomek, I. (1976a). *An Experiment with the Edited NearestNeighbor Rule*. IEEE Transactions on Systems, Man, and Cybernetics SMC-6, 448–452. https://doi.org/10.1109/TSMC.1976.4309523

Tomek I. (1976b). *Two modifications of CNN*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 6 (1), 769–772, 1976.

Usun, S. (2006). Uzaktan eğitim. *Ankara, Turkey: Nobel Publishing*.

Valentine, D. (2002). Distance learning: Promises, problems, and possibilities. *Online journal of distance learning administration*, *5*(3), 1-11.

Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, *9*.

Wang, D. (2023). Application of Machine Learning Technology in Classical Music Education. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, *18*(2), 1-15.

Web1. Students Adaptability Level in Online Education. https://www.kaggle.com/datasets/mdmahmudulhasansuzan/students-adaptability-level-in-online-education. . Access date February 09, 2023.

Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*, 408-421.

Yamamoto, G. T., & Altun, D. (2020). The Coronavirus and the rising of online education. *Journal of University Research*, *3*(1), 25-34.

Zhou, T., & Jiao, H. (2023). Exploration of the stacking ensemble machine learning algorithm for cheating detection in large-scale assessment. *Educational and Psychological Measurement*, *83*(4), 831-854.

Author Information

Ayşe Alkan

https://orcid.org/ 0000-0002-9125-1408

Samsun Science and Art Center

Turkiye

Contact e-mail: *ayse.alkan55@gmail.com*