sustainability

MDPI

Article

A Study in the Early Prediction of ICT Literacy Ratings Using Sustainability in Data Mining Techniques

Kyungyeul Kim [1], Han-Sung Kim [2], Jaekwoun Shim [3] and Ji Su Park [4,*]

[1] Department of Artificial Intelligence, Dongguk University, Seoul 04620, Korea; aooskrap6920@gmail.com
[2] Software Policy & Research Institute, Seongnam-si, Gyeonggi-do 13488, Korea; hansung@spri.kr
[3] Korea University Center for Gifted, Seoul 02841, Korea; jaekwoun.shim@gmail.com
[4] Department of Computer Science and Engineering, Jeonju University, Jeonju 55069, Korea
* Correspondence: jisupark@jj.ac.kr; Tel.: +82-63-220-2249

Abstract: It would be very beneficial to determine in advance whether a student is likely to succeed or fail within a particular learning area, and it is hypothesized that this can be accomplished by examining student patterns based on the data generated before the learning process begins. Therefore, this article examines the sustainability of data-mining techniques used to predict learning outcomes. Data regarding students' educational backgrounds and learning processes are analyzed by examining their learning patterns. When such achievement-level patterns are identified, teachers can provide the students with proactive feedback and guidance to help prevent failure. As a practical application, this study investigates students' perceptions of computer and internet use and predicts their levels of information and communication technology literacy in advance via sustainability-in-data-mining techniques. The technique employed herein applies OneR, J48, bagging, random forest, multilayer perceptron, and sequential minimal optimization (SMO) algorithms. The highest early prediction result of approximately 69% accuracy was yielded for the SMO algorithm when using 47 attributes. Overall, via data-mining techniques, these results will aid the identification of students facing risks early on during the learning process, as well as the creation of customized learning and educational strategies for each of these students.

Keywords: sustainability; data-mining techniques; early prediction of learning outcomes; information and communications technology literacy; education data mining

Citation: Kim, K.; Kim, H.-S.; Shim, J.; Park, J.S. A Study in the Early Prediction of ICT Literacy Ratings Using Sustainability in Data Mining Techniques. *Sustainability* **2021**, *13*, 2141. https://doi.org/10.3390/su13042141

Academic Editors: Jin Su Jeong and Danial Hooshyar
Received: 1 December 2020
Accepted: 15 February 2021
Published: 17 February 2021

1. Introduction

In learning scenarios, it is important for teachers to be able to identify students potentially at risk of faring poorly within a learning area and provide educational intervention proactively. Educational institutions are becoming increasingly concerned with achieving such interventions early on in the learning process [1] because estimating the ratio of positive-to-negative learning outcomes (i.e., succeeding or failing to learn) is critical to strategic planning. Through analysis of the variables from a student's background, it is possible to identify whether or not the student will be likely to succeed prior to immersion in the learning experience. Subsequently, appropriate actions can be taken to facilitate successful outcomes [2]. The capacity to analyze and predict academic performance represents an important milestone in the educational domain, and it is an important factor in building a student's future [3,4]. Therefore, these predictive variables can be used to identify students' learning characteristics to create adaptable methods for providing high-quality education to improve learning outcomes [5,6].

Existing studies have shown that students leverage relevant personal variables and attributes for their academic progress during instruction. These studies focused on the possibility of predicting academic achievement by utilizing student background factors as determined by surveys conducted prior to a class. Such background factors are necessary to analyze students' perceptive capabilities. If automated methods could be employed for

this purpose, the prediction of the learning outcomes could significantly reduce teacher assessment obligations [7,8]. Distinguishing student academic success or failure in advance can be accomplished by analyzing questionnaires created using sustainability-in-data-mining techniques. Such factors reflect personal, socioeconomic, psychological, and other diverse environmental variables.

The aim of this study is, therefore, to predict the academic performance of students using learning data obtained and analyzed via data-mining algorithms. Survey questions are created regarding the perception of the background factors related to information and communication technology (ICT) literacy, and the answers to the questionnaire are used to predict academic performance in advance. This study provides a model for enhancing the early prediction and awareness of the academic achievement of students. At an early stage in the learning process, our model can help identify students potentially at risk and provide quality intervention options that are appropriate for implementation in each case. Our method also aids in the provision of useful resources for creating customized learning and teaching strategies for these students.

This paper addresses three primary questions. First, are the data-mining techniques that utilize student perceptions, as mined from the questionnaire, effective in predicting early learning outcomes? Second, what data-mining techniques can predict student-learning outcomes? Finally, can we model the changing number of variables that reflect changes in the effectiveness of the data-mining techniques?

2. Application of Data-Mining Techniques in Education

Data-mining techniques are used to discover new information hidden within large databases [1,6]. Owing to advances in computing technology, these techniques are increasingly being used to solve problems and make discoveries in various fields of science, medicine, finance, and business [9,10]. In particular, data mining is being used in the field of education to diagnose students' learning factors and provide them with a variety of educational services [11,12].

The education industry leverages data-mining techniques to predict academic performance in advance lessons. The mined data relate to elements of the entire learning course (e.g., midterm, quiz, and activity content). Online and offline programming introductory courses applied similar metrics using neural-network (NN), DT, SVM, and NBC methods. These results showed that the SVM algorithms were the most efficient after 50% completion of a course. Failure rates were predicted with 92% efficiency in online classes and 83% efficiency in offline classes [13]. When predicting early time periods for majors in information-technology (IT)-related areas, seven algorithms (i.e., DT, rule induction, artificial NN, KNN, NBC, and random forest) were used. In this study, year-2007 student data were used for training, and the predicted rate was expressed using similar data from 2008. The results showed that NBC had a prediction rate of 83.7% [14]. University informatics courses used REPTree, J48, and M5P data-mining techniques to predict student performance. The attributes used to create the models included exam conditions, exam points, activities points, and more. The predictive model showed an average of 65% positive results and could reasonably predict a student's academic achievement [1].

However, for the early prediction of overall academic performance, graduation credits, or final-grade ratings, directly relevant attributes (e.g., exam and quiz scores) are commonly used. These related attributes are highly correlated with collected and predicted data, and assessments can be used to early-predict a student's achievement, but only after the learning process begins.

There are two ways to assess success or failure likelihood after the learning process begins. First, the research must ensure early prediction of the overall performance or required credits. Second, the student can express an early prediction rate based on the responses to a personal questionnaire provided prior to the learning process. For high-school students, a DT algorithm was used to predict student achievement, which was divided into five rating categories: "Unsatisfactory" (6%), "Basic" (40%), "Moderate" (38%),

"Good" (14%), and "Excellent" (3%). The data used included measures of self-esteem, self-concept, habits, motivation, cognitive skills, study strategies, and emotional variables representing personal factors related to academic performance. The prediction accuracy in that study was the highest in the "Basic" category with 40% of the student distribution. The remaining categories were in the range of 34–83% [15]. For college students, three algorithms (i.e., DT, NN, and SVM) were used to predict academic performance. The data included measures of online time, frequency of internet connection, amount of internet traffic, and usage behaviors online, which are linked to academic performance. The results showed that the SVM algorithm was the most accurate when predicting passing and failing grades (69–73%), followed by NN (68–71%) and DT (60–62%) [16]. The data used for college students included measures of age, gender, personality, motivation, and learning strategy, and data mining was used to predict the learning outcomes.

The results of that study indicated that SVM (73.3%) was the highest among the six algorithms, followed by KNN (69.4%), NN (69.0%), NBC (69.0%), DT (65.9%), and logistic regression (60.0%). Finally, for college students, the results were more accurate for freshmen than for seniors [17]. In the current paper, the early prediction of academic performance using extant learning processes is precluded, and the attributes directly relevant to predicting final grades are excluded. Additionally, the perception of IT-related students, which constitutes non-grade data focused on predicting final performance, is predicted using six sustainability-in-data-mining algorithms.

3. ICT Literacy

ICT literacy has been emphasized as an ability to be acquired by all to keep pace with IT development. Such literacy includes the ability to use digital technologies to solve problems, analyze, and generate information based on data, and communicate with others [18,19]. This is the interactions generated by learning to facilitate teacher decision-making when big data are generated, these big data are managed, and analyzed by data mining [20]. Since 2007, ICT literacy tests have been employed, and IT-related perceptions have been surveyed among elementary- and middle-school students in Korea [21,22].

The ICT literacy-test questionnaire comprises 36 questions concerning the internet, computer literacy, and IT curricula for daily life. The test results are divided into four levels (i.e., excellent, average, basic, and poor) according to student achievement. The criteria for each level are determined via expert consultation and consideration of the student's ability. The surveys of IT-related students measure the perceptions of their ability to use computers, smart devices, internet tools, and software. Details are shown in Table 1.

Table 1. ICT Measurement Elements.

No	Code Number	Survey Contents
1	grade	Elementary- and middle-school grades
2	gender	Gender
3	location	Place of residence
4	q_no1	Ability to identify the operating system being used
5	q_no2	Ability to use computer and internet for information search, music, video, blog, etc.
6	q_no3	Ability to use the computer's operating system
7	q_no4	Ability to manage smart devices by connecting them to computers
8	q_no5	Ability to solve errors related to information equipment
9	q_no6	Ability to use word-processing programs
10	q_no7	Ability to use spreadsheet programs
11	q_no8	Ability to use presentation programs
12	q_no9	Ability to use graphics programs
13	q_no10	Ability to use multimedia programs
14	q_no11	Ability to obtain, install, and use necessary programs
15	q_no12	Ability to install programs for multimedia playback, such as video
16	q_no13	Ability to download and print search documents from the internet
17	q_no14	Ability to upload materials to the internet

Table 1. *Cont.*

No	Code Number	Survey Contents
18	q_no15	Ability to attach and send materials by email
19	q_no16	Ability to register and communicate as a member on social media
20	q_no17	Ability to directly operate a simple notification service, mini homepage, and blog
21	q_no18	Ability to search for articles such as music, videos, and newspapers
22	q_no19	Ability to prevent computer viruses and malware
23	q_no20	Ability to protect information communication ethics, such as copyright
24	q_no21	Ability to prevent internet addiction
25	q_no22	Ability to cope with cybercrime and block spam mail
26	q_no23	Experience in joining the internet community
27	q_no24	Internet community operation experience
28	q_no25	Paths to learning how to use computers and the internet
29	q_no26	Ability to use computers and the internet (e.g., practice, review, educational, and information search)
30	q_no27	Computer or internet uses per week
31	q_no28	Degree to which computers (internet) should be taught at school
32	q_no29	Degree to which computers (internet) are used during the day to understand school homework or study
33	q_no30	Whether to use the internet when you want to use it (within 30 min)
34	q_no31	Whether to use the computer when you want to use it (within 30 min)
35	q_no32	Whether you have a desktop (desktop computer)
36	q_no33	Internet access at home (home)
37	q_no34	Whether you are studying on the computer or internet at school
38	q_no35	Whether or not you have wireless internet access capable mobile-phone information device
39	q_no36	When to use your computer for the first time
40	q_no37	Whether you have a computer with wireless internet access (e.g., laptop or desktop)
41	q_no38	The degree to which you use your computer (internet) throughout the day
42	q_no39	Whether the computer can be installed
43	q_no40	Whether you have a personal digital assistant information device capable of wireless internet access
44	q_no41	Whether you have a mobile phone (e.g., laptop or tablet)
45	q_no42	Programming ability (one or more languages, such as Java, C, AJAX, ASP, Visual Basic, PHP)
46	q_no43	Smart devices with wireless internet access (e.g., iPad)
47	q_no44	Smart devices (e.g., iPad, or Galaxy tablet)

4. Research Method

The proposed method predicted ICT literacy levels using sustainability-in-data-mining techniques based on students' IT-related perceptions. The ICT literacy rating prediction used six algorithms and data mining. A dataset from 2011 was used as the training set, and an attribute selector set of 47, 24, and 17 attributes were used for elementary schools, depending on the information gain ranking and empirical method. Similarly, sets of 47, 22, and 14 attributes were selected for middle-school students.

The data-mining technique selected six sets of algorithms referenced in the preceding studies. Several algorithms were used, including rule-based machine learning, OneR, DT, J48, ensemble listeners, bagging, random forest, neural networks, MLP, SVM, and sequential minimal optimization (SMO) [23]. This study used 10-fold cross-validation to create an optimal method for evaluating model performance [24].

The proposed model applied a dataset from 2012 as the test set, and the model's predictive accuracy was evaluated by measuring accuracy, precision, recall, and F1 score (i.e., F-measure). The flowchart of the ICT literacy evaluation prediction is shown in Figure 1.

4.1. Research Subject

The subjects of the study were selected by surveying students corresponding to 1% of the number of elementary- and middle-school students in Korea using stratified random sampling. For the 2011 dataset, 12,373 elementary-school students and 15,556 middle-school students were selected. Similarly, 12,905 elementary and 18,072 middle-school

students were selected for 2012. In total, 25,296 elementary- and 33,628 middle-school students participated in this study for two years.

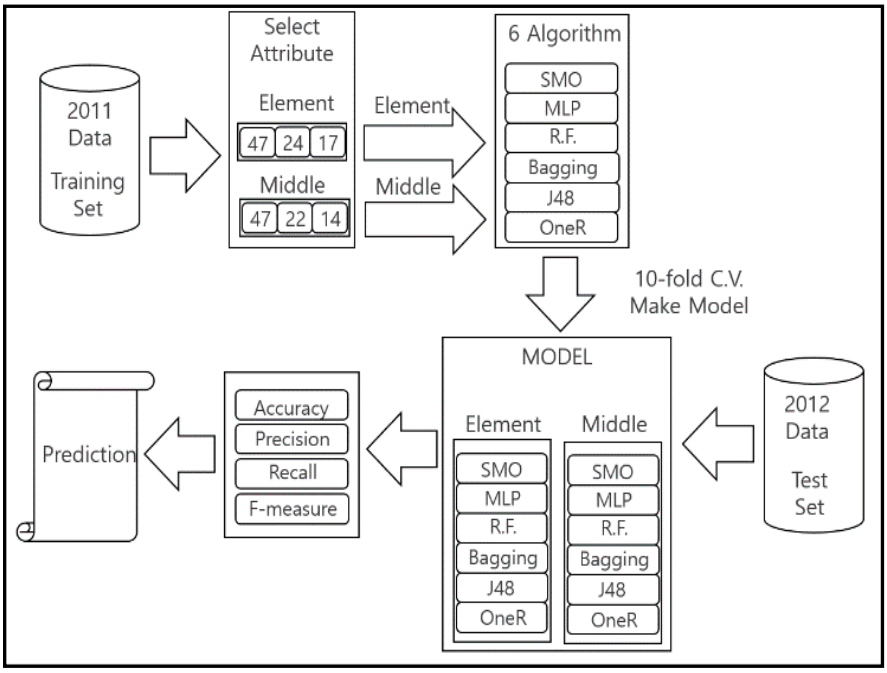

Figure 1. Flowchart for predicting student grades in ICT literacy.

4.2. Research Procedure

4.2.1. Preprocessing Data

ICT literacy results and IT questionnaire data of the elementary- and middle-school students were collected for the years 2011 and 2012. The data corresponding to 2011 were used as the training set, and those corresponding to 2012 were used as the test set. The criteria for data purification required the selection of missing values and excluded outliers that resulted in unstable or distorted data.

In this study, attribute selection was performed to improve the efficiency of data prediction [5]. The data attributes used in the analysis of data prediction were extracted from the 2011 training dataset using information gain and average merit. Information gain can determine the importance of a given attribute when deciding which attributes in the training dataset are most useful for distinguishing the classes to be learned, including the order of attributes. Using the heuristic method, attributes related to the research issues were included, and unnecessary items were deleted (e.g., user ID, student name, and registration number) to finally select the appropriate attributes. The details regarding this procedure are shown in Table 2.

4.2.2. ICT Literacy Class Classified By Experts

The ICT literacy grades were divided into four levels and were evaluated and classified by experts. For 2011, elementary schools had the highest percentage of average classifications: 19.8% "Excellent," 40.5% "Average," 36.5% "Basic," and 3.2% "Poor." The middle school had the highest percentage of basic classifications, with 6.3% "Excellent," 31.4% "Average," 57.8% "Basic," and 4.5% "Poor." In 2012, elementary schools had the highest percentage of basic classifications: 20.2% "Excellent," 30.0% "Average," 40.0% "Basic," and 6.8% "Poor." The middle schools had the highest percentage of basic classifications, with 9.7% "Excellent," 22.1% "Average," 60.4% "Basic," and 7.8% "Poor." The details regarding this are shown in Table 3.

Table 2. 2011 elementary- and middle-school information gain ranking attribute.

	2011 Elementary-School Ranked Attributes		2011 Middle-School Ranked Attributes	
No	**Average Merit**	**Attribute**	**Average Merit**	**Attribute**
1	0.129	q_15	0.095	q_14
2	0.104	q_13	0.092	q_05
3	0.101	q_16	0.085	q_03
4	0.096	q_14	0.067	q_06
5	0.093	q_18	0.066	q_08
6	0.085	q_17	0.065	q_4
7	0.084	grade	0.060	q_12
8	0.076	q_8	0.059	q_19
9	0.067	q_6	0.056	q_6
10	0.061	q_4	0.053	q_1
11	0.058	q_1	0.049	q_11
12	0.056	q_12	0.048	q_8
13	0.056	q_22	0.046	q_17
14	0.056	q_19	0.042	q_22
15	0.055	q_24	0.033	q_5
16	0.052	q_20	0.027	q_10
17	0.046	q_23	0.025	q_3
18	0.041	q_11	0.025	q_7
19	0.041	q_21	0.021	q_20
20	0.035	q_3	0.021	q_9
21	0.033	q_10	0.019	q_25
22	0.030	q_7	0.018	q_21
23	0.030	q_5	0.013	grade
24	0.030	q_9	0.013	q_24
25	0.023	q_3	0.013	q_39
26	0.020	q_26	0.012	q_3
27	0.017	gender	0.010	q_36
28	0.015	q_25	0.010	q_23
29	0.015	q_27	0.008	location
30	0.014	q_28	0.008	q_27
31	0.013	q_29	0.008	q_26
32	0.013	q_30	0.008	q_40
33	0.011	q_31	0.007	q_28
34	0.010	q_32	0.006	q_30
35	0.010	q_33	0.006	q_34
36	0.010	q_34	0.005	q_33
37	0.010	q_35	0.005	q_31
38	0.010	location	0.005	q_42
39	0.010	q_36	0.004	q_29
40	0.010	q_37	0.004	q_32
41	0.010	q_38	0.003	q_37
42	0.010	q_39	0.002	q_38
43	0.004	q_40	0.002	gender
44	0.003	q_41	0.002	q_35
45	0.003	q_42	0.001	q_41
46	0.001	q_43	0.000	q_44
47	0.001	q_44	0.000	q_43

4.2.3. Parameter Setting and Final Model Confirmation

The proposed ICT literacy rating prediction model increased the efficiency of the results when using the six data-mining algorithms. The analysis of the results for prediction was performed using 10-fold cross-validation to change the attributes of the data and basic option parameters and to adjust the highest prediction rate.

The proposed model used data mining to compare actual and predicted data results. As a result, models having higher accuracy were considered to be better.

Table 3. 2011–2012 ICT literacy grades.

Grade	Elementary School		Middle School	
	2011	**2012**	**2011**	**2012**
Excellent	2446 (19.8%)	2613 (20.2%)	985 (6.3%)	1747 (9.7%)
Average	5013 (40.5%)	4268 (33.0%)	4891 (31.4%)	3998 (22.1%)
Basic	4517 (36.5%)	5163 (40.0%)	8987 (57.8%)	10,917 (60.4%)
Poor	397 (3.2%)	879 (6.8%)	693 (4.5%)	1410 (7.8%)
Total	12,373 (100%)	12,905 (100%)	15,556 (100%)	18,072 (100%)

4.3. Data-Mining Techniques Used

Regarding the data-mining techniques, six algorithms were selected by comparing and analyzing their performance accuracies and capabilities based on previous studies. The OneR algorithm is a simple classification rule that is typically applied to a dataset to test a particular attribute. It is a simple and accurate classification algorithm that can create one rule for each predictor and select the rule having the smallest number of errors [25]. The J48 algorithm determines classification criteria based on normalized entropy difference and uses the concept of information entropy to create a DT from the learning data [26]. Bagging is used for statistical classification and regression, and it is an ensemble meta-algorithm designed to improve safety and accuracy. It can reduce the distribution of unstable procedures, such as regression trees, while greatly improving predictive accuracy [27]. Random forest is an ensemble learning method used for the creation, classification, and regression operations of multiple decision trees during training cycles. The benefits of random forest are that it selects one optimal solution, but it randomly selects from the *k* best options, thereby improving the decision trees [28]. The MLP is a kind of feed-forward artificial NN comprising at least three node hierarchies in which each node, except the input node, is a neuron that uses a nonlinear activation function [6]. The SMO algorithm is sensitive to fine-tuning, but manual fine-tuning is not desirable because it does not guarantee the efficiency of results [13].

4.4. Evaluation Criteria

In this study, accuracy, precision, recall (sensitivity), and F1 score were used as criteria for evaluating the six data-mining algorithms [29,30]. Accuracy is the percentage of the measurement that matches the actual and predicted values of the algorithm among the total data (1). Precision is the ratio between actually correct predictions of the positive class (true-positive (TP)) and all predictions of the positive class by the proposed model (TP + false positive (FP)). In other words, it is the ratio of what the algorithm predicted to be the correct answer (2). Recall (sensitivity) is the ratio of actual correct answers (TP + false negative (FN)) when the correct answer was accurately predicted (TP) (3). Precision and recall can be biased if there are many positives or negatives in the data, and the F1 score is used for the performance evaluation of the model using the harmonic mean of precision and recall (4).

$$Accuracy = \frac{\sum TP + \sum TN}{\sum Total\ population} \quad (1)$$

$$Precision = \frac{\sum TP}{\sum TP + \sum FP} \quad (2)$$

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \quad (3)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

5. Research Results

The proposed method predicted student ICT literacy levels using sustainability-in-data-mining techniques based on the perceptions of those IT-related learners. Information gain can be used to transform datasets to determine attribute importance and to distinguish classes. Therefore, the attributes found in the elementary-school results were divided into 47, 24, and 17 based on the average merit value of information gain ranking. The attributes from the middle-school results were divided into 47, 22, and 14.

The early-predicted results for elementary- and middle-school ICT literacy were characteristic of the algorithm used in the sustainability-in-data-mining techniques, indicating normal changes with the number of choices in the attributes.

5.1. ICT Literacy-Level Prediction Results for Elementary Students

The results of the ICT literacy-level predictions for the 2012 elementary-school dataset showed that the accuracy corresponded to the number of selected attributes. The lowest accuracy was 62.8%, and the highest was 67.3%. The highest early prediction result of all six algorithms was provided by SMO (67.3%), which used 47 attributes. The lowest prediction result was provided by OneR (62.8%), which used 17 attributes. The details regarding these results are shown in Figure 2.

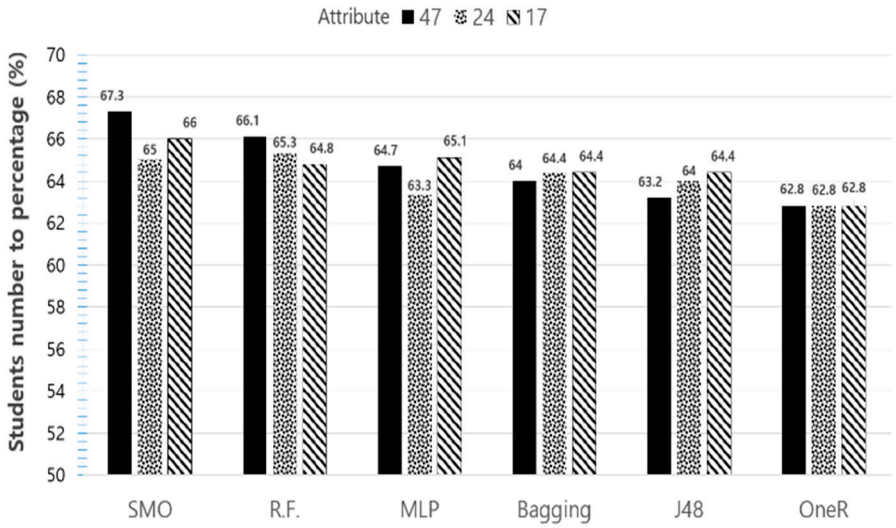

Figure 2. Accuracy based on the number of attributes (elementary school).

The F1 score uses the harmonic average of precision and recall and is an indicator of test and prediction. In this study, the SMO algorithm scored the highest (0.499) when 47 attributes were used. The lowest prediction result was returned by OneR (0.388) using 17 attributes. The details regarding these results are shown in Table 4.

Table 4. F1 score based on number of attributes.

Attribute	SMO	RF	MLP	Bagging	J48	OneR
47	0.499	0.478	0.455	0.453	0.438	0.388
24	0.467	0.469	0.438	0.457	0.450	0.388
17	0.480	0.462	0.462	0.456	0.455	0.388

5.2. Prediction Results for Middle-School Students

The results of the ICT literacy grade predictions using the 2012 IT-related middle-school dataset showed varying accuracies according to the number of selected attributes. For this dataset, the accuracy ranged from 63.9% to 68.7%. As noted, the highest prediction score was provided by SMO, which used 47 attributes (68.7%). This was also the highest

score achieved when comparing those of the other algorithms. The lowest prediction score was provided by MLP, which used 14 attributes (63.9%). The details regarding these results are shown in Figure 3.

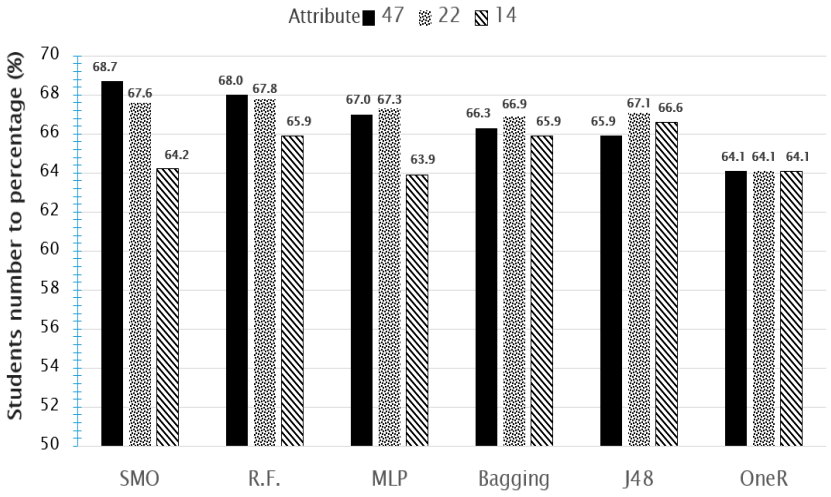

Figure 3. Accuracy based on the number of attributes (middle school).

The F1 score is an indicator of how well a prediction matches reality, and it uses a harmonic mean of precision and recall. As a result, when 47 attributes were used, SMO (0.541) exhibited the highest score. The lowest prediction score was provided by OneR (0.504) using 14 attributes. The details regarding these results are shown in Table 5.

Table 5. F1 score based on number of attributes.

Attribute	SMO	RF	MLP	Bagging	J48	OneR
47	0.541	0.531	0.534	0.520	0.526	0.504
22	0.525	0.535	0.531	0.525	0.525	0.504
14	0.504	0.519	0.505	0.517	0.524	0.504

6. Conclusions

This paper presented a model for predicting early academic performance-based learning perception using sustainability-in-data-mining techniques. Specifically, ICT literacy levels were predicted using six algorithms based on the students' perception of IT and ICT factors.

The highest SMO algorithm prediction results were 67.3% when using 47 attributes, and the lowest SMO algorithm prediction results were 65.0% when using 24 attributes for the 2012 elementary-school dataset. Therefore, the difference between the two cases was 2.3%. For the 2012 middle-school dataset, the highest and lowest prediction results for the SMO algorithms differed by approximately 4.5%, with accuracy scores of 68.7% for 47 attributes and 64.2% for 14 attributes.

The differences between early prediction results for the elementary- and middle-school datasets using the six-algorithm data-mining technique were 2.3% and 4.5%, respectively. By arranging the attributes affecting these results, similar scores can be achieved without significant changes in early prediction accuracy, even when a small number of features is selected. In particular, the accuracy results of the elementary- and middle-school students were more favorable when 24 and 17 attributes, respectively, were used than when all 47 were used. This was true for the MLP, bagging, and J48 algorithms. Therefore, the accuracy is dependent on both the characteristics of the algorithm and the number of attributes.

These results fully answer the three research questions presented in the introduction. The results of these three data-mining techniques can, therefore, be used to inform

teachers, institutions, and students in advance of potential learning successes or failures. Moreover, this innovation has the potential of avoiding or mitigating negative learning outcomes while providing students with important insights into improved educational approaches. In summary, it is possible to sufficiently predict early academic performance using sustainability-in-data-mining techniques based on student perceptions of IT competency and ICT literacy. Moreover, during the process of predicting early achievement by recognition, the SMO and RF algorithms were shown to be most effective. Finally, it was determined that the early prediction accuracy remained close to the highest observed ratio without significant changes when the number of attributes was reduced.

I examined the top five attributes of Information Gain ranking from the analyzed attributes. At the elementary school, the ability to attach data, download search documents, communicate on SNS, music, videos, and search articles using the internet was revealed. At middle school, computer virus prevention, the ability to use the internet, the ability to use the operating system, and the ability to resolve errors were shown. Middle school students used more specialized methods of using a computer than elementary school students. On the other hand, I examined the properties under Information Gain ranking. In general, we've found attributes that are not related to ICT, such as whether to keep smart devices or when to use a computer first boot. Analyzing the attributes indicated by this Information Gain, it can be said that they are related to pursuing the direction of learners' learning and educational strategies.

The significance of this study is its development of a new model for the early prediction of academic performance. This can help identify students facing risks early in the learning process via the application of data-mining techniques and the creation of customized learning and educational strategies for each student. Future research will require improvements to these study results via the extension and integration of the analysis of more diverse data to improve prediction accuracy.

This study has certain limitations, particularly, although the use of sustainability-in-data-mining techniques to predict achievement using student perception is interesting, it poses some risks. First, more data are required because, in this study, ICT literacy was only analyzed for 1% of Korea's student population via stratified random sampling. This is insufficient to represent larger populations. Second, only six representative algorithms (i.e., SMO, RF, MLP, bagging, J48, and OneR) were selected and studied. The addition of deep-learning algorithms, wherein sustainability-in-data-mining techniques are rapidly evolving, represents an important consideration for future work.

Author Contributions: Conceptualization, K.K. and J.S.; Data curation, K.K. and J.S.; Formal analysis, K.K.; Investigation, J.S.; Methodology, K.K.; Project administration, J.S.P.; Resources, H.-S.K.; Software, H.-S.K.; Supervision, J.S.P.; Validation, J.S.P.; Writing—original draft, K.K.; Writing—review & editing, H.-S.K., J.S. and J.S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Natek, S.; Zwilling, M. Student data mining solution-knowledge management system related to higher education institutions. *Expert Syst. Appl.* **2014**, *41*, 6400–6407. [CrossRef]
2. Romero, C.; Ventura, S.; Pechenizkiy, M.; Baker, R.S. (Eds.) *Handbook of Educational Data Mining*; CRC Press: Minneapolis, MN, USA, 2010; pp. 57–58.
3. Kaur, P.; Singh, M.; Josan, G.S. Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Comput. Sci.* **2015**, *57*, 500–508. [CrossRef]

4. Naseer, M.; Zhang, W.; Zhu, W. Early Prediction of a team performance in the initial assessment phases of a software project for sustainable software engineering education. *Sustainability* **2020**, *12*, 4663. [CrossRef]
5. Sorour, S.E.; Goda, K.; Mine, T. Comment data mining to estimate student performance considering consecutive lessons. *J. Educ. Technol. Soc.* **2017**, *20*, 73–86.
6. Ian, H.W.; Eibe, F.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: San Mateo, CA, USA, 2016.
7. Xing, W.; Guo, R.; Petakovic, E.; Goggins, S. Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Comput. Hum. Behav.* **2015**, *47*, 168–181. [CrossRef]
8. Hinojo-Lucena, F.J.; Aznar-Díaz, I.; Cáceres-Reche, M.P.; Trujillo-Torres, J.M.; Romero-Rodríguez, J.M. Factors influencing the development of digital competence in teachers: Analysis of the teaching staff of permanent education centres. *IEEE Access* **2019**, *7*, 178744–178752. [CrossRef]
9. Liao, S.H.; Chu, P.H.; Hsiao, P.Y. Data mining techniques and applications–A decade review from 2000 to 2011. *Expert Syst. Appl.* **2012**, *39*, 11303–11311. [CrossRef]
10. Kim, B.; Kim, J.; Yi, G. Analysis of clustering evaluation considering features of item response data using data mining technique for setting cut-off scores. *Symmetry* **2017**, *9*, 62. [CrossRef]
11. Romero, C.; Ventura, S. Educational data mining: A review of the state of the art. *IEEE Trans. Syst. Man Cybern. C* **2010**, *40*, 601–618. [CrossRef]
12. Agaoglu, M. Predicting instructor performance using data mining techniques in higher education. *IEEE Access* **2016**, *4*, 2379–2387. [CrossRef]
13. Costa, E.B.; Fonseca, B.; Santana, M.A.; de Araújo, F.F.; Rego, J. Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Comput. Hum. Behav.* **2017**, *73*, 247–256. [CrossRef]
14. Asif, R.; Merceron, A.; Ali, S.A.; Haider, N.G. Analyzing undergraduate students' performance using educational data mining. *Comput. Educ.* **2017**, *113*, 177–194. [CrossRef]
15. Martinez Abad, F.; Chaparro Caso López, A.A. Data-mining techniques in detecting factors linked to academic achievement. *School Effectiv. Sch. Improv.* **2017**, *28*, 39–55. [CrossRef]
16. Xu, X.; Wang, J.; Peng, H.; Wu, R. Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Comput. Human Behav.* **2019**, *98*, 166–173. [CrossRef]
17. Gray, G.; McGuinness, C.; Owende, P. An application of classification models to predict learner progression in tertiary education. In Proceedings of the 2014 IEEE International Advance Computing Conference (IACC), Gurgaon, India, 21–22 February 2014; pp. 549–554.
18. Lee, S.; Kim, J.; Lee, W. Analysis of elementary students' ICT literacy and their self-evaluation according to their residential environments. *Indian J. Sci. Tech.* **2015**, *8*, 81–88. [CrossRef]
19. Siddiq, F.; Hatlevik, O.E.; Olsen, R.V.; Throndsen, I.; Scherer, R. Taking a future perspective by learning from the past -A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educ. Res. Rev.* **2016**, *19*, 58–84. [CrossRef]
20. Pozo-Sánchez, S.; López-Belmonte, J.; Rodríguez-García, A.M.; López-Núñez, J.A. Teachers' digital competence in using and analytically managing information in flipped learning. *Cult. Educ.* **2020**, *32*, 213–241. [CrossRef]
21. KERIS. *Assessing Student's ICT Literacy at a National Level*; Korea Education & Research Information Service: Daegu, Korea, 2011.
22. Kim, J.; Lee, W. Meanings of criteria and norms: Analyses and comparisons of ICT literacy competencies of middle school students. *Comput. Educ.* **2013**, *64*, 81–94. [CrossRef]
23. Malhotra, R. A systematic review of machine learning techniques for software fault prediction. *Appl. Soft Comput.* **2015**, *27*, 504–518. [CrossRef]
24. Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Stat. Comput.* **2011**, *21*, 137–146. [CrossRef]
25. Holte, R.C. Very simple classification rules perform well on most commonly used datasets. *Mach. Learn.* **1993**, *11*, 63–91. [CrossRef]
26. Quinlan, R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers: San Mateo, CA, USA, 1993.
27. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
28. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Yin, C.; Zhou, B.; Yin, Z.; Wang, J. Local privacy protection classification based on human-centric computing. *Hum. Cent. Comput. Inform. Sci.* **2019**, *9*, 33. [CrossRef]
30. Ghrabat, M.J.J.; Ma, G.; Maolood, I.Y.; Alresheedi, S.S.; Abduljabbar, Z.A. An effective image retrieval based on optimized genetic algorithm utilized a novel SVM-based convolutional neural network classifier. *Hum. Cent. Comput. Inform. Sci.* **2019**, *9*, 31. [CrossRef]