



Article

Towards Independent Students' Activities, Online Environment and Learning Performance: An Investigation through Synthetic Data and Artificial Neural Networks

Malinka Ivanova ^{1,*} and Tsvetelina Petrova ²

- ¹ Department of Informatics, Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Blvd. Kl. Ohridski 8, 1797 Sofia, Bulgaria
- ² Department of Energy and Mechanical Engineering, Technical College of Sofia, Technical University of Sofia, Blvd. Kl. Ohridski 8, 1797 Sofia, Bulgaria; tsvetelina.petrova@tu-sofia.bg
- * Correspondence: m_ivanova@tu-sofia.bg

Abstract: During the pandemic, universities were forced to convert their educational process online. Students had to adapt to new educational conditions and the proposed online environment. Now, we are back to the traditional blended learning environment and wish to understand the students' attitudes and perceptions of online learning, knowing that they are able to compare blended and online modes. The aim of this paper is to present the performed predictive analysis regarding the students' online learning performance taking into account their opinion. The predictive models are created through a supervised machine learning algorithm based on Artificial Neural Networks and are characterized with high accuracy. The analysis is based on generated synthetic datasets, ensuring a high level of students' privacy preservation.

Keywords: independent student activities; intelligent online environment; learning performance; machine learning; artificial neural networks; predictive analysis; synthetic data; privacy preservation



Citation: Ivanova, M.; Petrova, T. Towards Independent Students' Activities, Online Environment and Learning Performance: An Investigation through Synthetic Data and Artificial Neural Networks. *Informatics* **2023**, *10*, 37. <https://doi.org/10.3390/informatics10020037>

Academic Editor: Sotiris Kotsiantis

Received: 16 February 2023

Revised: 5 April 2023

Accepted: 6 April 2023

Published: 10 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During the last few years, online education has become more popular and useful worldwide because of the pandemic situation [1,2]. The pandemic has forced most educational institutions to replace face-to-face classes and blended learning forms with fully online education [3]. In such conditions, the crucial role in organizing a high-quality education has been played by the educational software environment with its functional and technical parameters. Contemporary learning environments possess intelligent tools for content management and delivery, communication, and assessment, including tools for learning analytics. In a systematic review regarding learning analytics in online environments prepared by Kew and Tasir [4], they found that the current learning analytics tools are mainly focused on students' activities monitoring, extracting statistical analysis, making predictions, and, in some cases, performing acts in the form of interventions.

The students accept learning analytics because it provides support and facilitation of the learning process, guiding them through different learning paths to achieve successful final results. On the other hand, the students are concerned about the privacy of their data and the way it is used. Therefore, they require high demands not only on a technologically oriented and high-quality learning process, but also that their data are protected and secure [5]. One contemporary data privacy technology, which is applied in this work to ensure the educational privacy of the students, is based on the generation of synthetic data. Synthetic data is used to protect persons whose data is collected, as the extracted statistics are nearly similar to those gathered from the original dataset [6]. Additionally, this approach is preferable because the created machine learning models are based on the generated synthetic datasets, and, in this way, they are protected against attacks for revealing students' identities or sensitive information and against malicious data mining.

Artificial intelligence and machine learning techniques are grasped as sophisticated and effective instruments for the analysis and prediction of multiple variables in an educational process, including student learning performance [7,8]. Artificial Neural Networks (ANNs) as a part of machine learning are also utilized as a base for modeling and forecasting topics in Higher Education, mining students' data and proposing adaptive learning models [9]. Many researchers are looking for the right predictors/factors influencing the performance of students in order to prognosis and prevent some negative activities, events, or processes that could trouble or lead to failure of the participants in an educational process.

The main goal of this paper is to demonstrate an approach for predictive analysis of student online learning performance with care for privacy preservation based on synthetically generated datasets and utilization of ANNs. Additionally, this paper presents the results from evaluating the accuracy of the created machine learning models. The contributions can be summarized as follows:

1. Proposition of a methodology based on synthetic data generation for analyzing the students' opinions regarding the role of the online environment in completing independent activities;
2. Creation of a model, based on ANNs, for predicting the online learning performance of students and evaluation of its accuracy;
3. Evaluation of the quality of the synthetically generated datasets for proving the principle of a balance between privacy preservation and data usefulness.

The paper is organized as follows: Section 2 discusses the role of online learning environments in supporting the students' effectiveness according to recently published scientific papers. Section 3 summarizes the application of ANNs for analyzing and predicting the student learning performance. Section 4 explains the proposed methodology. Section 5 includes the students' opinions regarding the benefits of an online environment. In Section 6, the independent activities and concerns of the students are discussed, considering the answers of surveyed participants. Section 7 summarizes the students' opinions regarding their effectiveness in an online environment. Analytical prediction of the student learning performance is presented in Section 8. Section 9 includes our conclusions.

2. Online Learning Environment and Students' Effectiveness

In the literature, there are different definitions of the term "students' effectiveness", considering a wide variety of factors and the context of the application. Thus, this definition could be described in more general or more specific aspects, as it is noted in the explored papers below. In this work, when the term "students' effectiveness" in the online environment is utilized by authors, its meaning is related to a process of conducting an independent learning activity for a given period with a certain level of quality. The students' learning effectiveness influences their learning performance.

There are a lot of currently published scientific works outlining the benefits and bottlenecks related to students' effectiveness and online environment usage. This research is carried out with respect to the online environments in developing the students' knowledge, facilitating a learning process and improving learning performance through analysis.

Gudkova et al. [10] investigate the effectiveness of the learning process using Moodle in the student's independent work in terms of online learning of English. The respondents are teachers and students. Both of them assess learning and teaching the English language on the Moodle platform as easier, really efficient, and more productive, but also with more advantages than disadvantages. The findings show that Moodle helps students to revise and better understand the material studied in class. The authors conclude that Moodle is an effective tool for organizing, maintaining, and assessing the independent work of undergraduate students. The main detected problems are the lack of computer literacy and IT skills for both teachers and students and some technical issues—access to the Internet, stability of the connection, etc.

Umek et al. [11] examine the impact of the Moodle platform on students' performance. In the conducted study, students from the University of Ljubljana, Faculty of Public

Administration, participated in the period from 2008 to 2014. The results show that the implementation of a Moodle-based e-learning system leads to a statistically significant increase in students' performance, measured as the average grade and the average number of admissions to the exams.

Rajan and Manyala [12] assess the students' perception of e-learning and the use and effectiveness of Moodle. The investigation includes 263 students registered for the Introductory Physics course. They give their opinion through an online survey. The authors conclude that the students still do not use the e-learning system effectively because they do not have enough knowledge about Moodle's tools and access to the platform because of a lack of internet. The results show that the students mainly use the Moodle platform for downloading the materials (81%) and a small portion of them for delivery of assignments (13%). Only 5% use Moodle to communicate with lecturers, and just 1% receive announcements from the lecturers.

Wang et al. [2] investigated 519 students from a medical university in order to examine the connection between e-learning self-efficacy, monitoring, willpower, attitude, motivation, strategy, and the e-learning effectiveness of college students in the context of online education during the outbreak of COVID-19. The authors' findings are that e-learning motivation strongly and positively affects e-learning effectiveness. Students' e-learning experience is an important factor in learning efficiency. There are differences in learning effectiveness between students with and without e-learning experience. Therefore, the conclusion of the authors is that students should obtain initial knowledge about online learning platforms before starting with online education.

Dyrek et al. [13] investigate the perception of e-learning among 615 students of Medical Universities in Poland. Respondents have to answer questions about the perception of lecturers' effectiveness, assessment of stationary and online classes, as well as changes in learning habits and restrictions on education. Findings show that students highly appreciate online lectures and seminars in contrast to laboratory and clinical classes. A number of the students have positive attitudes toward e-learning. The main pointed problems of e-learning are the quality of the learning materials (26.9%), restrictions in direct contact with the teacher (19.6%), Internet connection (16.8%), and home conditions (13.8%).

Encarnacion et al. [14] examine the effectiveness and impact of e-learning on the teaching and learning process at Oman Tourism College in Muscat, Sultanate of Oman. For that purpose, 60 teachers and 162 students took part in the survey. The authors conducted a study that measures the effectiveness of e-learning using the following five factors: content quality, assessment, collaborative environment, system quality, and technical support. The authors find out that teachers and students appraise the e-learning platform (Moodle) as an effective tool to increase the delivery of instruction and develop knowledge acquisition skills through the transfer of learning. Moreover, e-learning education improves collaboration. On one hand, the teachers feel a very positive impact on their working styles by using e-learning as part of face-to-face instruction. On the other hand, the students think they become motivated to learn independently and study their courses with bigger responsibility.

Tan and Chen [15] monitor and examine synchronous online learning courses in Physics at the Singapore University of Technology and Design. They observe the learning process for 13 weeks, during which they have to manage 3 main challenges—(1) how to replicate the collaborative learning within the online classroom; (2) how to conduct physics demonstrations; and (3) how to give non-verbal feedback and maintain student engagement online. The authors conclude that the conducted synchronous course is successful. Moreover, they think about shifting/implementing some of the used educational technologies from this synchronous course in both the online and the classically performed classrooms.

Tan et al. [16] present the results of their pilot study about virtual product dissection used in blended synchronous learning. For that purpose, 19 students were involved in an elective energy systems course. The authors' findings are that this learning mode is

suitable, the students can learn as effectively even though they are not on campus, and it is worth thinking about how to extend this type of available online content.

A summary of explored literature sources is presented in Table 1, which outlines some important functions of contemporary online environments for facilitating the organization of students' independent work, for announcements delivery, communication, collaboration, assessment, and feedback. The positive attitude to e-learning is also highlighted, and the importance of online environments for better understanding of the class material, achieving effective student learning, and improving learning performance is drawn. Among the bottlenecks pointed out are a lack of needed computer skills, poor knowledge of Moodle tools, lack of contact with the teacher and content quality, and emerging technical problems regarding Internet connection.

Table 1. Summary of literature sources regarding the role of the online environment and some bottlenecks.

| Literature Source | The Role of an Online Environment | Bottlenecks |
|-------------------------|---|--|
| Gudkova et al. [10] | <ul style="list-style-type: none"> Facilitate understanding of the class material; Support organization, maintenance, and assessment of the independent work. | <ul style="list-style-type: none"> Lack of computer literacy and IT skills; Technical issues related to Internet connection. |
| Umek et al. [11] | <ul style="list-style-type: none"> Significant increase in the student performance | |
| Rajan and Manyala [12] | <ul style="list-style-type: none"> Support downloading the materials; Assessment delivery; Communication; Announcement receiving. | <ul style="list-style-type: none"> Lack of enough knowledge about Moodle tools; Lack of Internet. |
| Wang et al. [2] | <ul style="list-style-type: none"> Positive effect on e-learning effectiveness | |
| Dyrek et al. [13] | <ul style="list-style-type: none"> Support online lectures and seminars; Positive attitude to e-learning. | <ul style="list-style-type: none"> Problems with content quality, Internet connection, and contact with the teacher. |
| Encarnacion et al. [14] | <ul style="list-style-type: none"> Effective tool for instruction delivery and knowledge development; Improved collaboration; Motivation for independent learning and bigger responsibility. | |
| Tan and Chen [15] | <ul style="list-style-type: none"> Support collaborative learning, online demonstrations, non-verbal feedback, and student engagement. | |
| Tan et al. [16] | <ul style="list-style-type: none"> Support effective student learning. | |

3. Application of Artificial Neural Networks in the Analysis of Students' Learning Performance

This section presents some applications of ANNs for analyzing students' learning performance and outlines the main factors/predictors influencing the performance of learning activities.

The architectures of ANNs are inspired by the workability of human neural networks in the performance of some tasks [17]. They are constructed as connected neurons, which form the following layers: input, output, and hidden. The neurons transmit or not the input signal with a given weight to other parts of the ANN as it is modeled through several activation functions. For solving a concrete task, a specific ANN must be developed by specifying the number of input and output variables, the number of hidden layers, the number of neurons at each layer, and the type of activation function. Although ANNs are far from the perfection of the human brain, they are often used to solve simple tasks with high accuracy. ANNs are currently the preferred way to represent a number of problems due to their architectural flexibility and easy parameterization.

Dharmasaroja and Kingkaew [18] build ANN models for predicting students' learning performance as the main considered predictors are students' learning background, average grade points, and demographic issues. The models based on ANNs and logistic regression are compared as the authors see the big potential of ANNs for learning performance prediction.

Huang et al. [19] propose a two-step model based on a support vector machine and ANN algorithms for predicting the students' course performance as well as proving its effectiveness through experimentation. Among the considered important factors for measuring students' learning performance are geographical, demographic, and social parameters and type of learning activities for better engagement.

According to Hamadneh et al. [20], the four main factors influencing the students' learning performance in a blended learning environment are attendance, face-to-face or virtual studying, scores on the midterm exam, and percentage of performed assessments. Additionally, they propose an ANN-based model to predict the learning performance of students, decrease students' failure, and improve the learning process overall.

Leelaluk et al. examine the relationship between the reading behavior of the students and their performance [21]. The authors claim that by taking into account the data of the students' reading activities and considering the content of the lecture, it can be predicted whether the student is at risk or not. Reading behavior is also further explored when changing the sequence of lectures in the course. Predicting at-risk students can give the instructors a clue that these students need assistance and support.

Academic performance is investigated by Ahmad and Shahzadi, and main predictors are considered: time for studying, academic interaction, habits for studying, skills for learning, hardworking, the students' marks before, CGPA (Cumulative Grade Point Average) of the second semester and environment at home [22]. With high accuracy, the created ANN can predict the students at risk.

This literature review shows the potential of ANNs to make analysis and predictions towards the students' learning performance considering different factors: from students' background to students' behavior and conducted learning activities in an online environment. The considered factors/predictors and utilized machine learning techniques are summarized in Table 2.

Table 2. Predictors for predicting students' learning performance.

| Literature Source | Predictors | Machine Learning Techniques |
|--------------------------------|--|---|
| Dharmasaroja and Kingkaew [18] | <ul style="list-style-type: none"> • Students' learning background; • Average grade points; • Demographic issues. | Comparison between ANNs and logistic regression |

Table 2. Cont.

| Literature Source | Predictors | Machine Learning Techniques |
|-------------------------|---|---|
| Huang et al. [19] | <ul style="list-style-type: none"> • Geographical; • Demographic; • Social parameters; • Type of learning activities for better engagement. | A two-step model based on support vector machine and ANN algorithms |
| Hamadneh et al. [20] | <ul style="list-style-type: none"> • Attendance; • Face-to-face or virtual studying; • Scores on the midterm exam; • Percentage of performed assessments. | ANN-based model |
| Leelaluk et al. [21] | <ul style="list-style-type: none"> • Reading behavior of the students | Multilayer perceptron |
| Ahmad and Shahzadi [22] | <ul style="list-style-type: none"> • Time for studying; • Academic interaction; • Habits for studying; • Skills for learning; • Hardworking; • The previous students' marks; • CGPA of the second semester and • Environment at home. | Multilayer perceptron |

4. Proposed Methodology

The aim of the proposed methodology is to present a safety predictive analysis of processing synthetic students' data and creating ANN-based models for prognosis of the learning performance in online environment. Firstly, a survey tool is designed with three groups of questions: (1) the importance of the online environment for learning performance, (2) accomplishment of online independent activities and students' concerns, (3) online environment utilization and students' effectiveness. The survey mainly includes questions with several possible answers, and the results exceed 100%. There is no possibility for free responses. The survey is online and delivered to students at Technical University of Sofia enrolled in different courses in the domains of engineering and informatics. It is based on anonymity of students' responses and is driven by voluntary participation. The number of responders is 223 students.

The methodology consists of the following procedures and is shown in Figure 1.

Procedure 1: Data collection through online distributed online survey and pre-processing for meeting the requirements regarding the file format (.csv) and machine learning utilization is conducted. For each question, a dataset is prepared. A part of the dataset with randomly selected records for the first question regarding the students' attitudes to online education is presented in Table 3. NA indicates that the student did not select this option in the answer.



Figure 1. Proposed methodology for safe machine learning processing.

Table 3. A part of the dataset for the first question.

| Save Resources | Save Health | Online Education Is Suitable for Some Courses | Prefer Blended Education | Prefer Online | Prefer Face-to-Face |
|----------------|-------------|---|--------------------------|---------------|---------------------|
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| saveResources | NA | onlineSuitabilitySomeCorses | NA | preferOnline | NA |
| NA | saveHealth | NA | NA | NA | NA |
| saveResources | NA | NA | NA | preferOnline | NA |
| ... | ... | ... | ... | ... | ... |

Procedure 2: The originally pre-processed datasets are used for smart generating synthetic datasets through the online platform MOSTLY AI [23], which is GDPR-driven. A synthetic dataset is created for each question. A part of the synthetic dataset for the

first question is presented in Table 4. Evaluation of the quality of the generated synthetic datasets is also conducted by considering parameters such as distribution and accuracy (Figure 2). The accuracy matrix shows the similarity between the original (target) and synthetic data. The darker color presents bigger similarity, while the lighter color indicates bigger difference. The distribution parameter Distance to Closest Record (DCR) highlights the closeness between the original (target) and synthetically generated datasets. For good privacy preservation, the distribution of the synthetic data should be close but not the same as the distribution of the original data. Here, the principle of a balance between good privacy preservation and useful statistics is considered.

Table 4. Synthetically generated dataset for the first question.

| Save Resources | Save Health | Online Education Is Suitable for Some Courses | Prefer Blended Education | Prefer Online | Prefer Face-to-Face |
|----------------|-------------|---|--------------------------|---------------|---------------------|
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| saveResources | NA | NA | NA | preferOnline | NA |
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| saveResources | saveHealth | NA | NA | preferOnline | NA |
| ... | ... | ... | ... | ... | ... |

Q1. What is your attitude toward the online educational environment?

Privacy

Accuracy: 97.8%



Figure 2. Evaluation of privacy preservation for the first question.

Procedure 3: Synthetic datasets processing in RapidMiner Studio platform [24], applying a deep learning algorithm, is performed. The created predictive model consists of 4 ANNs as the architecture of each ANNs includes an input layer with 3 or 4 variables, 3 different hidden layers, respectively, with 50, 60, and 70 neurons in each layer and an output layer with 1 variable. The first ANN predicts the importance of the online learning environment for the student’s learning performance, the second ANN predicts the students’ concern at working with the online environment, the third model predicts the students’ online effectiveness, and the fourth ANN prognosis the students’ online learning performance. Evaluation of the accuracy of the constructed predictive model is also conducted. The datasets for learning by the ANNs are created after summarizing the datasets of all questions according to their categorization in groups. A part of the created dataset regarding the first group of questions considering the importance of online environment

for students learning performance is shown in Table 5. Similar datasets are created for the rest two groups of questions.

Table 5. Collected dataset for the first group of questions.

| Learn Knowledge That Is Not Directly Related to Major Topic | Consider Advantages of Online Learning | Difficulties with Online Learning | Consider Importance of a Virtual Room | Possibility for Projects Defense | Possibility for Online Examination | Consider Communication Tools | Importance |
|---|--|-----------------------------------|---------------------------------------|----------------------------------|------------------------------------|------------------------------|--------------------|
| NA | yes | NA | yes | NA | yes | yes | Important |
| NA | yes | NA | yes | yes | yes | yes | Important |
| yes | NA | yes | NA | NA | yes | NA | Somewhat Important |
| ... | ... | ... | ... | ... | ... | ... | ... |

5. Online Learning Environment and Its Benefits

Analysis of responses of the surveyed students is conducted on the generated synthetic data to save their privacy and feelings for a safe educational process. The analysis is performed by taking into account the questions grouped into the following three different groups: (1) the importance of the online environment for learning performance, (2) the accomplishment of online independent activities and students’ concerns, and (3) the online environment and students’ effectiveness.

The aim of the first group of questions is to reveal the students’ attitudes regarding the utilization of an online learning environment and its capability for organizing and delivering effective and high-quality learning. It is worth mentioning that some of our students are working professionals aged over 19 years.

Q1: The first question reveals the students’ attitude towards online education. The question allows more than one possible answer. A total of 83.4% of the students like online educational forms because it saves time and resources for traveling. A total of 58.3% of the students also like online education because it is an opportunity for their health saving and saving the health of their families (during the pandemic). Additionally, 29.1% said online education is suitable only for some courses, but others must be performed in specialized laboratories and face-to-face halls. A total of 21.9% prefer a blended educational environment as a combination of physical presence and online participation. Alternatively, 51.56% said that they prefer face-to-face education in comparison to online one. The students’ answers show that a big group of them possess a positive attitude towards the online educational environment, benefiting from its advantages. Another not-so-small number of students express their preferences to be educated in a face-to-face educational mode, which is understandable due to the nature of the studied courses: engineering and informatics.

Q2: The students’ answers confirm that utilization of an online environment opens up different opportunities, as 62.78% of them share that during online learning, they learn a lot of new things that are not directly related to their major but are useful and practical in terms of their (future) work/hobbies. University classes organized through an electronic environment receive positive feedback from the students—results show that 81.61% of the interviewees start looking for online courses related to their (future) work and/or hobbies after seeing the advantages of online learning. A small part of the students, 3.13%, said that online learning is difficult for them because there are too many differences in the way classes are conducted compared to face-to-face learning.

Q3: To the question, “Is it a challenge for you to learn through an online learning system?”, 61.43% of the students answered that it is not challenging to learn with the eLearning system; rather, it is easy, and the necessary knowledge is acquired quickly and imperceptibly. 13.9% of them also do not think that it is a challenge because they have a lot of experience with learning in an online environment. However, 8.96% of them answer

this question with “yes, it is a challenge” because they do not possess such competencies to take advantage of the online environment and gain a rich learning experience.

Q4: In the answers to the question, “What made online learning difficult for you?”, a very big part of the students, 80.71%, responded that they do not experience any difficulties. The rest of them share several emerging problems with camera/microphone/router, internet connection, the lack of separate room at home, laptop/smartphone/tablet, and they have to invest more resources in new hardware devices or for improving the internet connection.

Q5: An important question is “Which capabilities/tools of the online learning environment are useful and important for you to ensure an effective learning process?” A high vote of students (74.43% of them) is given to the software tool for screen sharing and its capability for the performance of real-time consultations/receiving help/obtaining a piece of advice. A total of 77.57% express the importance of the environment for conducting online assessment tasks and online examinations. Moreover, 73.54% of them like the possibility of using various visual materials (presentations and videos) in online classes. A total of 78.2% of the students highlight the importance of the availability of a virtual room for the conductance of synchronous educational tasks. Additionally, 67.71% of them like the possibility of online submission and online defense of course projects and course works. Over half the students, 52.46%, evaluate the advantages of synchronous and asynchronous communication tools in the online environment and the possibility of receiving messages from other students and educators.

This analysis reveals the students’ positive attitude to the online environment and its functionality because it proposes possibilities for gaining new knowledge, saving time and additional resources, and their health (in the pandemic). A big part of the students like and appreciate the benefits of the online environment for learning material delivery and presentation, the conductance of synchronous and asynchronous educational tasks, and examination. Furthermore, a small number of students express their concern about working with the tools of the online environment, as well as having some difficulties from a technical point of view.

The evaluation of the produced synthetic data shows that they have high quality according to accuracy, which is between 96% and 97.9%, and the DCR distributions between the original (target) and synthetic data are close enough (Figure 3).

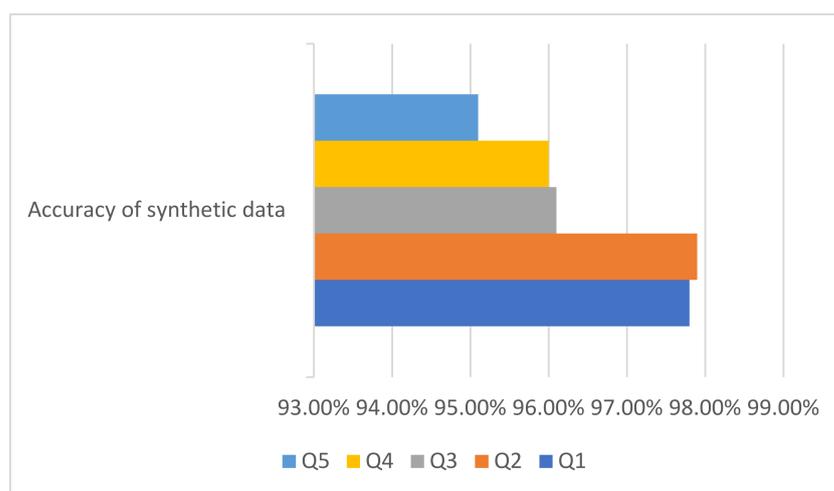


Figure 3. Evaluation of privacy preservation for the first group of questions.

6. Independent Activities and Concerns of the Students

The purpose of this group of questions is to understand the students’ attitude to conduct independent activities and the supportive role of the online learning environment.

Q6: The question “What is your opinion about the assignments that you have to prepare independently (course work/course project/presentation/essay/programming code)?” gives various perspectives regarding the independent students’ activities. A total of 60.73% of them respond that their attitude is positive because they provoke creativity/curiosity and give freedom/scope for expression. Nearly half, 48.85%, of them give a positive answer for another reason because they have to study further and thus acquire new knowledge and skills useful for (future) work/hobby. Additionally, 49.77% share that their positive opinion is based on the possibility to combine and apply knowledge and skills in practical tasks, i.e., how to develop themselves and with each subsequent task to be performed better and better. A total of 53.88% of responses is also related to a positive answer because along with the specific material for which they are doing research and learning something new, they have to use the obtained knowledge and skills so far: formatting, data processing, preparing presentations in a visual form (graphs and tables), etc. On the other hand, 2.28% of the students share their negativism concerning difficulty related to learning a lot of new things to complete a learning activity. A total of 7.76% said that it makes it difficult for them, and sometimes, they have to ask colleagues for help to complete a task.

Q7: The aim of the next question is to understand the meaning of independent students’ work in a learning process: “What arouses your interest in the study course to a greater extent?” A total of 47.08% of the students answered that their interest in a given course increases when complete information during lectures is given, additional tasks are assigned for better understanding, and the learning material is consolidated during exercises. Moreover, 40.8% of the students said that their interest arises when full and comprehensive information on the studied subject during lectures is given. A total of 10.76% responded that only introductory information during lectures and setting independent tasks were to look up additional information and then present it appropriately to remain interested in this course.

Q8: In order to understand the students’ preferences to perform learning activities, the following question is asked: “How do you prefer to work on assignments? (alone independently or in group)”. A total of 40.35% of the students prefer to work in a group because that is how they give more ideas and complement each other. On the other hand, 16.16% of them share their preferences to work in their own way independently because they have an opportunity to expose information in a form which they consider the most suitable. Moreover, 43.49% of the students answer that it depends on the type of learning activity.

Q9: To the question, “What type of independent task provokes your creativity and imagination in relation to the course studied (possibility of more than one answer)?”, the following answers are achieved: 52.01% of the students like to prepare presentations on a given topic, and it is related to their creative competences; 51.56% of them are creative at conductance of course work/project; 27.8% at essay writing; 23.31% at programming; 15.24% of the student are creative at any performance of any task, and its type does not matter; 14.34% of the students feel creative at infographics creation; 1.34% of the students feel comfortable at problems conceptualizing through concept maps.

Q10: According to the usefulness of independent tasks for personal development, the next question is: “Do you like the conductance of independent activities during your learning?” Over half the students, 52.01%, like it because it is the way to better understand the lecture material and laboratory practices. A total of 23.76% also like to perform independent activities because, nowadays, they could be completed in an easier manner, considering multiple information sources. The rest of the surveyed students do not like it because the independent activities are so time-consuming.

Q11: For developing independent work (abstract, presentation, term paper, course project, and programming code), 79.83% of the students are looking for several information sources, comparing information, and using more visual materials. On the other hand, 16.59% rely only on the information in Wikipedia, while 18.38% of the students use only

recommended by the educator relevant literature. Only 8.07% prefer to study a single information source and use it for the preparation of this independent task.

Q12: The next question, “When preparing your independent work (paper/presentation/term paper/term project) what is the most important?” is related to the quality of the result and learning performance. The majority of students, 74.88%, examine their work carefully, checking whether it meets the requirements and then submitting it. A total of 18.83% of them just check whether respond to the requirements, and 5.38% of the students prefer to write it quickly in order to have more free time.

From the students’ answers, it can be said that most appreciate and like the conductance of independent tasks because it develops their creativity, gives them an opportunity freely to express their point of view, and leads to obtaining new knowledge and skills. In preparing the final result of the independent tasks, many students answer that they prefer to investigate several information sources, not only those given by the educator, and they care about the high quality of their task performance. A very small portion of the students prepares their assignments quickly because they think they should not take much time to complete them.

The evaluated quality of synthetically generated data is presented in Figure 4, as the obtained accuracy is between 95.9% and 98.2%.

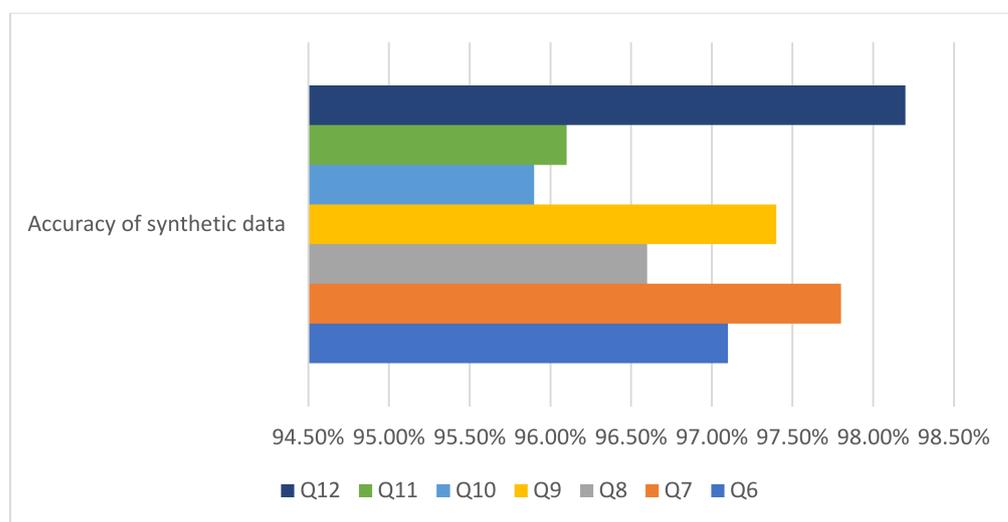


Figure 4. Evaluation of privacy preservation of the second group of questions.

7. Learning Effectiveness in Online Environment

Q13: The students are asked, “Does the online learning environment support your learning effectiveness?”, and 91.47% of them give a positive answer.

Q14: The next question reveals when the students’ learning performance is better—at fully online learning, face-to-face learning, or a blended form of learning. A total of 67.71% of the students share that their learning performance is better with fully online learning, 25.11% with blended learning, and the rest of the students answer face-to-face classroom.

Q15: On the question, “How do you describe your own concentration during online classes?”, 57.39% of the students answer that their concentration is constant regardless of the educational form (online or face-to-face), 28.69% of them share that they lose concentration more easily in face-to-face classes, while 13.92% of the students get distracted more often in online classes.

Q16: 52.01% of the students feel more comfortable asking/answering a question when they participate in online classes. For 43.04% of the students, the educational form does not matter—online, blended, or face-to-face. The rest of the students prefer to communicate in face-to-face classes.

Q17: According to the assessment form, 38.56% of the students prefer online exams, 34.97% of them like online assessments during the semester, and 18.38% of the students prefer the assessment to be based on a combination of the online exam and online independent assessment tasks during the semester. A very small portion of the students expresses their preference for face-to-face and paper-based exams.

Regarding the effectiveness of the students in an online environment, almost all of them respond positively and share that their learning performance is better. A very small number of students are not satisfied with online education and prefer all activities to be performed face-to-face.

Figure 5 summarizes the privacy and accuracy parameters of the synthetic data, and it can be seen that the balance principle is also respected, e.g., the synthetic data are close enough to the original (target). The accuracy of the produced synthetics is high: between 95.7% and 98.7%.

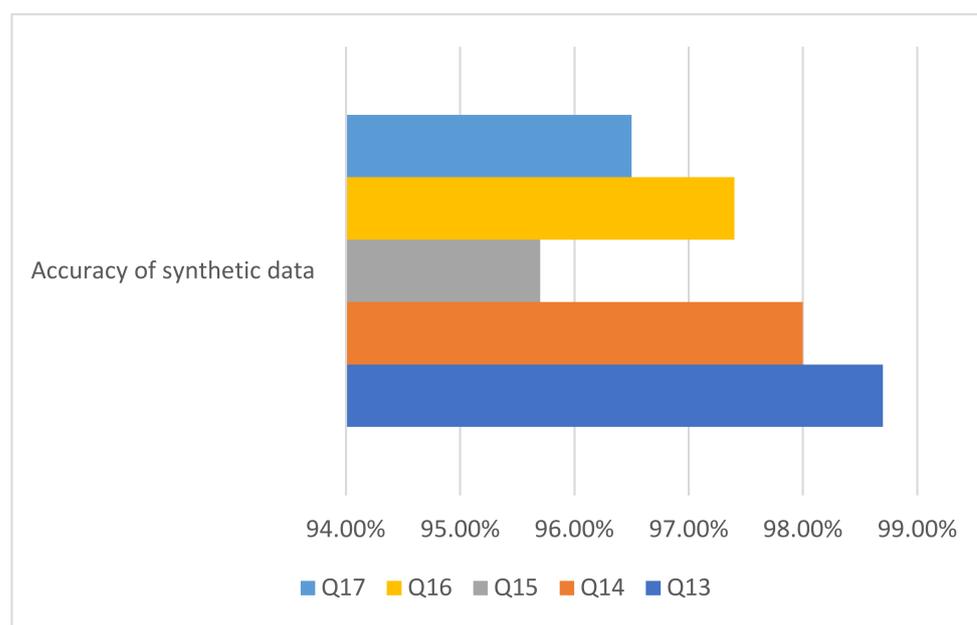


Figure 5. Evaluation of privacy preservation of the third group of questions.

8. Analytical Prediction of the Students' Online Learning Performance

To predict the students' learning performance, an ANN-based model on synthetic datasets is created, as presented in Figure 6. The model consists of four ANNs and their purpose is as follows.

The first ANN makes predictions regarding the importance of the online learning environment for student learning performance according to their opinion. The input variables are as follows: (1) opportunities in the online environment for learning support, (2) challenging issues in online learning for students, and (3) useful tools of the online environment which can facilitate learning. The output is the predicted importance of the online educational environment to assist the students' online learning activities. The predicted values can be important, somehow important, and unimportant.

The second ANN predicts the students' concerns regarding the conducted individual activities (including the individual activities for accomplishing a group task) in the online learning environment. The input variables are as follows: (1) students' attitudes at conductance of independent activities, (2) students' preferences on how to perform the individual activity (alone, in a group, or it depends on the activity type), (3) precision at preparing the product of an independent activity (examination and checking of the product, examination, checking, or fast preparation), and (4) type of the independent activity (presentation, essay, infographics, course project, concept map, and programming code), which leads to a

creative mood upon solution/product creation. The output variable predicts the students' concern regarding the quality of the performed independent activity as its values could be neural, concerned, or very concerned.

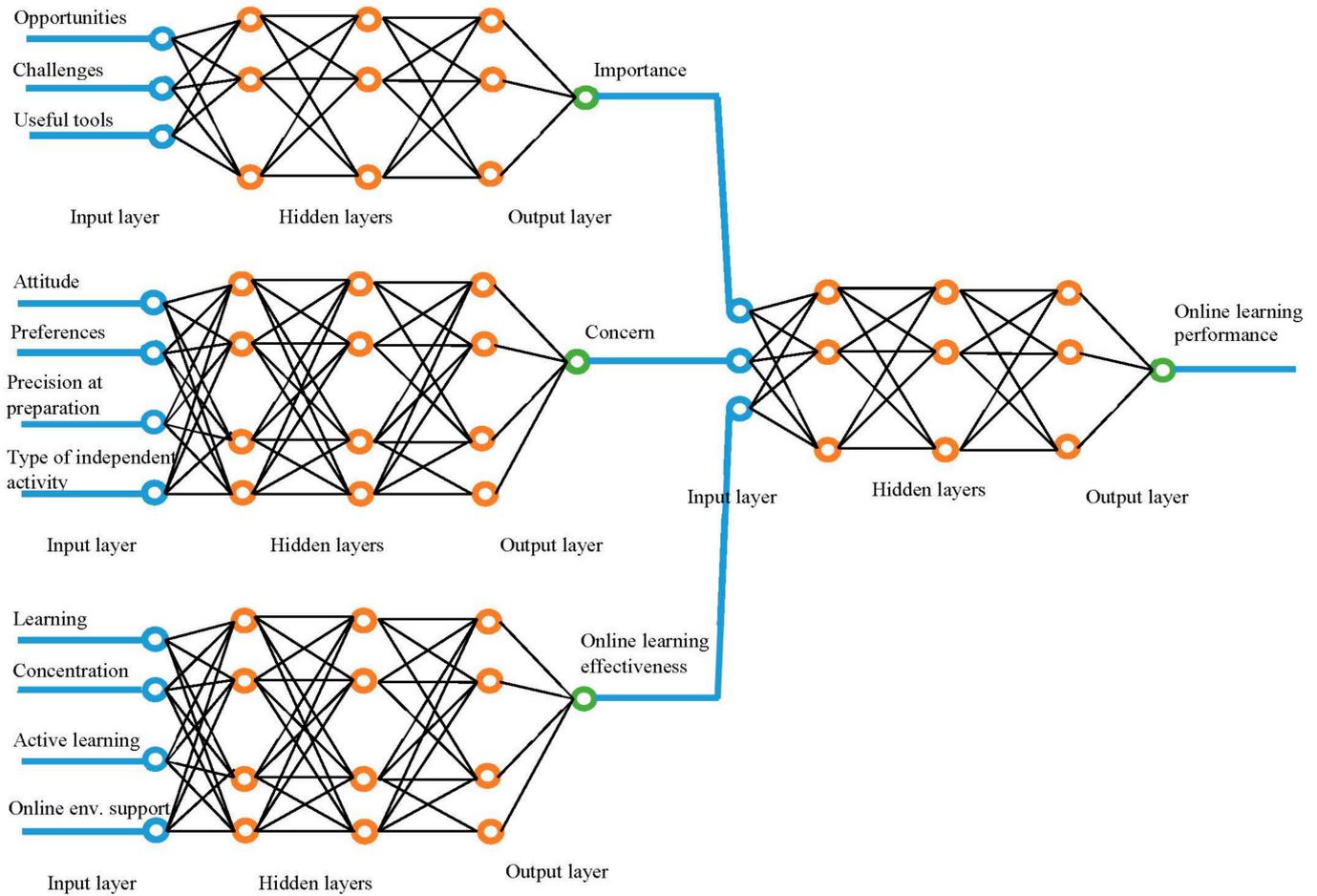


Figure 6. The created predictive model.

The third ANN prognosis of the students' online effectiveness according to the four inputs: (1) learning mode, which depends on when the student feels more effective (at online, blended, or face-to-face learning), (2) students' concentration in different educational forms (in online or face-to-face classes), (3) when the active learning is more provocative for a discussion (in online classes, at face-to-face classes, and the educational settings does not matter), and (4) whether the online environment provides the needed support in order to students to be effective learners. The predicted output is the students' online learning effectiveness, which could be acceptable, good, and very good.

The fourth ANN takes the inputs of the described above three ANNs, including importance, concern, and online learning effectiveness, to predict the students' online learning performance as the possible values are high, good, and acceptable. The predictive accuracy of the ANNs is evaluated as the performance vectors are summarized in Table 6.

Table 6. Evaluation of ANNs' performance vector.

| Performance Vector | |
|--|---|
| <i>First ANN with output importance</i> | <i>Second ANN with output concern</i> |
| Accuracy: 86.57% | Accuracy: 93.85% |
| Absolute_error: 0.2720 | Absolute_error: 0.1973 |
| Relative_error: 27.20% | Relative_error: 19.73% |
| Root_mean_squared_error: 0.3770 | Root_mean_squared_error: 0.2749 |
| <i>Third ANN with output Online Learning Effectiveness</i> | <i>Fourth ANN with output Online Learning Performance</i> |
| Accuracy: 98.51% | Accuracy: 92.42% |
| Absolute_error: 0.0528 | Absolute_error: 0.0979 |
| Relative_error: 5.28% | Relative_error: 9.79% |
| Root_mean_squared_error: 0.1351 | Root_mean_squared_error: 0.2311 |

9. Conclusions

This paper presents an approach for predictive analysis of students' online performance. It uses synthetic data, artificially generated, accounting for the embedded statistics in the original datasets. The obtained synthetic data is evaluated concerning the accuracy matrix and statistical distribution. The results outline the application of the principle of balance between data privacy and useful output. The synthetic datasets are then used for creating ANN-based models to predict the online learning performance of the students. The main considered predictors are the importance of the online environment for conducting the students' activities, concerns regarding the quality of the performed independent activity and online learning effectiveness when they must complete learning activities in the online environment. The created ANN models are characterized by high-performance vectors and with high accuracy to predict the students' online learning performance.

A limitation of this study is the considered data, which was taken during one semester from volunteer students. For this reason, in the future, a created predictive model should be verified with other data and checked if it is applicable to be refined and modified.

Author Contributions: Conceptualization: M.I. and T.P.; methodology: M.I.; formal analysis and investigation: T.P. and M.I.; resources: M.I. and T.P.; writing—original draft preparation: T.P. and M.I.; writing—review and editing: M.I. and T.P.; visualization: M.I.; funding acquisition: T.P. and M.I. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support. This research is supported by the Bulgarian FNI fund through the project, "Modeling and Research of Intelligent Educational Systems and Sensor Networks (ISOSeM)", contract KII-06-H47/4 from 26 November 2020.

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

- Holzer, J.; Lüftenegger, M.; Korlat, S.; Pelikan, E.; Salmela-Aro, K.; Spiel, C.; Schober, B. Higher Education in Times of COVID-19: University Students' Basic Need Satisfaction, Self-Regulated Learning, and Well-Being. *AERA Open* **2021**, *7*, 23328584211003164. [[CrossRef](#)] [[PubMed](#)]
- Wang, C.-Y.; Zhang, Y.-Y.; Chen, S.-C. The Empirical Study of College Students' E-Learning Effectiveness and Its Antecedents Toward the COVID-19 Epidemic Environment. *Front. Psychol. Sec. Educ. Psychol.* **2021**, *12*, 573590. [[CrossRef](#)] [[PubMed](#)]
- Alkabaa, A.S. Effectiveness of using E-learning systems during COVID-19 in Saudi Arabia: Experiences and perceptions analysis of engineering students. *Educ. Inf. Technol.* **2022**, *27*, 10625–10645. [[CrossRef](#)] [[PubMed](#)]
- Kew, S.N.; Tasir, Z. Learning Analytics in Online Learning Environment: A Systematic Review on the Focuses and the Types of Student-Related Analytics Data. *Technol. Knowl. Learn.* **2022**, *27*, 405–427. [[CrossRef](#)]

5. Liu, B.; Lu, J.; Wang, P.; Zhang, J.; Zeng, D.; Qian, Z.; Ge, S. Privacy-Preserving Student Learning with Differentially Private Data-Free Distillation. In Proceedings of the 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China, 26–28 September 2022; pp. 1–6. [CrossRef]
6. Wagner, P. Privacy Enhancing Technologies and Synthetic Data. 2020. Available online: <https://ssrn.com/abstract=3762686> (accessed on 17 November 2022). [CrossRef]
7. Wang, S.; Sun, Z.; Chen, Y. Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Educ. Inf. Technol.* **2022**, 23328584211003164. [CrossRef]
8. Tarik, A.; Aissa, H.; Yousef, F. Artificial Intelligence and Machine Learning to Predict Student Performance during the COVID-19. *Procedia Comput. Sci.* **2021**, *184*, 835–840. [CrossRef] [PubMed]
9. Okewu, E.; Adewole, P.; Misra, S.; Maskeliunas, R.; Damasevicius, R. Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review. *Appl. Artif. Intell.* **2021**, *35*, 983–1021. [CrossRef]
10. Gudkova, Y.; Reznikova, S.; Samoletova, M.; Sytnikova, E. Effectiveness of Moodle in student's independent work. In Proceedings of the E3S Web of Conferences, Rostov-on-Don, Russia, 24–26 February 2021; Volume 273. [CrossRef]
11. Umek, L.; Keržič, D.; Aristovnik, A.; Tomažević, N. An assessment of the effectiveness of Moodle e-learning system for undergraduate public administration education. *Int. J. Innov. Learn.* **2017**, *21*, 165–177. [CrossRef]
12. Rajan, R.; Manyala, R.O. Effectiveness of Moodle in the Learning of Introductory Physics During COVID-19 Pandemic: A Case Study at the University of Zambia. *Int. J. Innov. Sci. Res. Technol.* **2021**, *6*, 1124–1131. Available online: <https://ijisrt.com/assets/upload/files/IJISRT21FEB326.pdf> (accessed on 17 November 2022).
13. Dyrek, N.; Wikarek, A.; Niemiec, M.; Owczarek, A.J.; Olszanecka-Glinianowicz, M.; Kocełak, P. The perception of e-learning during the SARS-CoV-2 pandemic by students of medical universities in Poland—A survey-based study. *BMC Med. Educ.* **2022**, *22*, 529. [CrossRef] [PubMed]
14. Encarnacion, R.F.E.; Galang, A.A.D.; Hallar, B.J.A. The Impact and Effectiveness of E-Learning on Teaching and Learning. *Int. J. Comput. Sci. Res.* **2020**, *5*, 383–397. [CrossRef]
15. Tan, D.Y.; Chen, J.-M. Bringing Physical Physics Classroom Online—Challenges of Online Teaching in the New Normal. *Phys. Teach.* **2021**, *59*, 410. [CrossRef]
16. Tan, D.Y.; Kwan, W.L.; Koh, L.L.A.; Pee, G.-Y.M.; Lur, K.T.; Yeo, Z.Y. Virtual Dissection Activities as a Strategy for Blended Synchronous Learning in the New Normal. In Proceedings of the 2022 IEEE Global Engineering Education Conference (EDUCON), Tunis, Tunisia, 28–31 March 2022; pp. 565–570. [CrossRef]
17. Madhwaran, M.; Louzazni, M. Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications. *J. Electr. Comput. Eng.* **2022**, *2022*, 5416722. [CrossRef]
18. Dharmasaroja, P.; Kingkaew, N. Application of artificial neural networks for prediction of learning performances. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 745–751. [CrossRef]
19. Huang, C.; Zhou, J.; Chen, J.; Yang, J. A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction. *Neural Comput. Appl.* **2021**, 1–13. [CrossRef]
20. Hamadneh, N.N.; Atawneh, S.; Khan, W.A.; Almejalli, K.A.; Alhomoud, A. Using Artificial Intelligence to Predict Students' Academic Performance in Blended Learning. *Sustainability* **2022**, *14*, 11642. [CrossRef]
21. Leelaluk, S.; Minematsu, T.; Taniguchi, Y.; Okubo, F.; Shimada, A. Predicting student performance based on Lecture Materials data using Neural Network Models. In Proceedings of the CEUR Workshop 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior, Online, 21–22 March 2022; Volume 3120, pp. 11–20. Available online: <https://sites.google.com/view/lak22datachallenge> (accessed on 17 November 2022).
22. Ahmad, Z.; Shahzadi, E. Prediction of Students' Academic Performance using Artificial Neural Network. *Bull. Educ. Res. Dec.* **2018**, *40*, 157–164.
23. MOSTLY AI. Available online: <https://mostly.ai/synthetic-data/> (accessed on 23 January 2023).
24. RapidMiner Studio Manual. Available online: <https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf> (accessed on 23 January 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.