Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-20014-3

Application of the digital twin model in higher education

Aybeyan Selim[1] · Ilker Ali[1] · Muzafer Saracevic[1,2] · Blagoj Ristevski[3]

Received: 30 April 2023 / Revised: 10 January 2024 / Accepted: 2 August 2024

Abstract
Globalization and technological development are the two main pillars of the basis of modern enterprises. Different business decisions want to determine the balance between the potential of the company and the demand for its products or services. Making decisions requires timely and accurate information obtained from management information systems. The technology of digital twins is a relatively new technology, which proves its value day by day through optimization during the development of projects. This study presents the application model of digital twin technology, which uses the NFC (near-field communication) protocol for user authentication and authorization. As a bridge between the natural and digital, we use the digital twin to create a copy of the services in a digital environment. NFC technology dramatically simplifies the interaction between mobile accessories and allows data transfer between devices. Our model collects the data and allows storing, analyzing, and using them for different purposes. We tested the proposed model with the decision tree and the k-means algorithms. The testing results showed that digital twins are fruitful and promising perspectives in effective future educational process planning.

Keywords Digital twin · NFC technology · Data mining · Decision tree · Education

✉ Aybeyan Selim
aybeyan@vision.edu.mk

Ilker Ali
ilker@vision.edu.mk

Muzafer Saracevic
muzafers@uninp.edu.rs

Blagoj Ristevski
blagoj.ristevski@uklo.edu.mk

[1] Faculty of Engineering and Architecture, International Vision University, Gostivar, North Macedonia

[2] Department of Computer Sciences, University of Novi Pazar, Novi Pazar, Serbia

[3] Faculty of Information and Communication Technologies – Bitola, University "St. Kliment Ohridski", Bitola, Republic of Macedonia

Published online: 22 August 2024

1 Introduction

In recent years, digital data, accessed by many users from different environments, changed when necessary, and recorded in various forms, has become increasingly widespread due to these advantages [1].

Industry 4.0 have based on technology that enables the connection between the real and digital world [2]. The basic paradigms we associate with this term are the Internet of Things (IoT), Cloud computing, Digital-physical systems, and Big Data. The Internet of Things is a network of accessories connected via the Internet, enabling data exchange between them [3]. Connected devices are equipped with various electronic components, sensors, and appropriate computer support, including multiple programs and applications, and collect and exchange data [4]. The connection enables the interaction of devices and systems, facilitating their control and monitoring and opening space for new services [5]. Communication is achieved between machines, devices, and people, and digital-physical systems include both.

Digital twins are virtual replicants of physical products that model actual product behavior in real-time and serve to detect possible flaws and problems [6]. According to M. Greaves, the model of the digital twin concept includes three main components: physical products in real space, virtual products in virtual space, and data and information links that connect virtual and real things. The products' virtual representations perform various analyses, test new features, and simulate possible scenarios that would be expensive or even impossible to implement on a physical product [7]. With the help of sensors and numerous technologies in the digital environment, they display the elements and dynamics of products connected via the IoT, which allows monitoring of the device's state and territory and how they change [8]. The role of the sensor is to obtain correct information for stored data in the computer about the current state of the device and the environment [9]. In addition to physical data sources, computers also store data created in a digital environment, such as design specifications, lists of components and materials, calculation results, simulations, and the like [10]. Combining this data with data collected via sensors provides a complete picture of the system's operation and performance, allowing convenient operation in the phase of use and maintenance and better adaptation of the product to actual operating conditions [11] [12].

The digital twin's ultimate purpose is to reduce overall costs, optimize maintenance, create new products and services opportunities, and improve existing outcomes by optimizing components or reducing waiting times between operations connecting various subsystems [13].

Data collection begins digitally using various non-destructive technologies such as sensors, measuring devices, vision systems, etc. These technologies enable the connection of physical and virtual products, thus forming a complete digital-physical system consisting of a physical product in real space, a virtual model in virtual space, and data flow from physical to virtual environment [14].

The realm of education has witnessed a transformative shift with the integration of digital twin technology [15]. While the discourse around digital twins in education is gradually gaining traction, it is essential to acknowledge the impact of underlying technologies like Near Field Communication (NFC). Notably, NFC can be classified into two primary types: Passive NFC and Active NFC [16].

Passive NFC devices harness the electromagnetic fields generated by active devices, such as smartphones or NFC-enabled readers, and do not require an independent power source. Conversely, Active NFC devices are equipped with their own power source, allowing for more complex operations and extended communication ranges.

In the context of our educational framework, the choice between Passive and Active NFC is pivotal. Factors such as power consumption, communication range, and the specific requirements of our educational applications will guide this decision-making process. Considering the collaborative and dynamic nature of contemporary education, a careful evaluation of which type aligns better with our objectives is imperative [17].

Moreover, NFC stands out among touchless communication methods due to its distinct advantages. In contrast to other touchless methods, NFC offers enhanced security, efficiency, and versatility. The ability of NFC to provide seamless, short-range communication aligns with the active learning experiences that education strives to foster.

As we embrace the digital future of education, it's worth noting that almost all mobile devices today are equipped with NFC technology, opening up countless possibilities for innovative educational applications. In comparison to traditional touchless methods, NFC's ease of use, reliability, and widespread integration make it a compelling choice for modern educational environments [18].

1.1 Related works and previous applications of digital twins technologies

The number of references to digital twins technology in Education is relatively tiny. The literature on digital twins in Education is still scarce and does not shine with thematic depth and diversity. However, many experts in Education and science today understand and explain using these digital tools. We need to understand that the application of digital technology is in no way aimed at belittling and minimizing human potential in Education and deserves special attention and support [19]. Our fears that robots will replace us do not hold and must be emphasized that the future of Education lies in digitalization. The importance of virtual technologies is seen in helping students have an active experience rather than a passive learning experience and improving their creativity [20]. Thanks to virtual technologies, a positive education environment is created [21]. Developing technologies and virtual tools cause a significant change in education methods, including engineering education. In Education and other professional practices, more and more group work is practiced daily than traditional individual theory-based lectures [22]. Apart from that, project-based learning refers to learning from a specific engineering project as a case study, and there are many application examples [23]. Examples of group-based learning methods in engineering, problem-solving with open-ended solutions, hands-on projects, and team-oriented communication are also available [24]. The concept of learning and active participation of the student in the lesson is provided by digital technologies. With the help of these technologies, students learn more through active practice and are better prepared for their careers [24]. The research of Schuster et al. [25], sees that using new digital technologies such as the digital twin in Education affects students' employment and the companies' competitiveness.

Australian University researcher S. Sepasgozar refers to digital twin technology for online learning in architecture. The author notes that online learning represents five new digital technologies that use virtual and augmented reality in this study. A. Liljaniemi and H. Paavilainen, analyzing digital twin technology in engineering education, note that by introducing new digital technologies, such as digital twins, we can present new knowledge to students, teachers, and companies. They obtain that technology can increase learning motivation. Digital twin technology is also a tool for teaching students to work with production systems, and this is a justified proposal by introducing learning theory within pedagogical digital twins [26].

The most crucial educational reality content will shortly be setting and solving society and education tasks for synchronization and digital data replication on all objects and connections, properties, relationships, and regularities of the virtual, natural and social trench. One of the brightest signs of the new educational reality is the "digital" technology used in conjunction with the digital twin, gaining more and more independent popularity and stability in various industries. One of the brightest signs of the new educational reality is the "digital" technology used in conjunction with the digital twin, gaining more and more independent popularity and stability in various industries. In the paper of Selimi et al. [27], results demonstrated that the positive effect on teaching and the increase in knowledge mediated by interactive teaching tools and digital tools positively impact Education and interactively increase knowledge. Digital tools fully fulfill the tasks of the "digital agenda" and social and technological development in the process of a qualitatively new integration of society and Education, the transition to a global and cross-border level [28].

Digital twins are based on numerous, cumulative, real-world data collected from various sources, measured in real-time. This data creates a dynamic model of a product or process in digital form, which can provide a better insight into system performance and thus encourage the changes that need to be made to the physical system to improve it. Thanks to the rapid development of IT in recent years, the performance of computer systems has become higher while reducing their cost. Thanks to the rapid development of IT in recent years, the performance of computer systems has become higher while reducing their cost. It has become possible to perform complex algorithms with large amounts of data for various analyzes and simulations [29].

In the current landscape, nearly all mobile devices come equipped with Near Field Communication (NFC) technology, unlocking a myriad of possibilities for its integration into educational settings.

This potential is exemplified in the study presented by [30], where the application of NFC takes a tangible form in supporting the teaching of histology to university students. Described as HistoNFC, this innovative approach not only incorporates NFC technology but also integrates seamlessly with existing teaching methodologies. By combining various teaching proposals, including portals of contents and virtual microscopes embedded in mobile devices, HistoNFC allows students to access information at their convenience, offering flexibility in both time and location. Moreover, the technology enables educators to assess and evaluate students' activities in real-time, fostering a personalized learning experience.

The practical application of NFC in HistoNFC showcases its adaptability and efficacy in modern educational contexts. This novel approach represents a paradigm shift, providing a versatile tool that not only facilitates traditional teaching methods but also enhances the accessibility and personalization of educational content. As we explore the dynamic intersection of NFC technology and education, it becomes clear that such innovations hold the potential to reshape the landscape of teaching and learning [31].

1.2 Analytical foundations and dataset overview

This research is conducted at the International Vision University in North Macedonia, where the developed model has been applied to 652 students across five faculties. All students utilized the model for accessing course materials from lectures and exercise sessions.

The decision trees, a widely employed data mining approach for classification and prediction, were employed alongside the k-means algorithm to answer research questions and evaluate the application. The dataset incorporates final exam results from 2020 (face-to-face) and 2021 (online), containing various student records from the University Information System (UIS).

1.3 Statement of the problem

Our research focuses on exploring and analyzing the factors influencing students' academic success in online learning environments. As educational institutions increasingly transition to online platforms, understanding the dynamics that contribute to or hinder student success becomes imperative. The problem at hand is to identify the key variables and patterns that impact students' performance in exams, particularly during the shift from traditional face-to-face instruction to online modalities.

1.4 Challenges

Several challenges are encountered in our study, such as the absence of a one-size-fits-all approach to online learning success and the varied adaptability of students to technology. Additionally, global events like the COVID-19 pandemic introduce new challenges to the educational landscape, necessitating an examination of their impact on academic outcomes.

1.5 Goals of the study

1.5.1 Identification of key predictors

This study investigates crucial factors influencing students' success in online learning environments, considering variables like faculty, gender, and online activity.

1.6 Effectiveness of the model

We constructed predictive models using decision trees (CHAID algorithm) and k-means clustering to provide actionable insights for educational practitioners.

1.7 Evaluation the impact of online learning

Assessed the impact of the transition from face-to-face to online learning on student success by comparing academic performance before and during the COVID-19 pandemic.

1.8 Feature extraction

In our study, feature extraction is crucial for identifying and utilizing relevant information from the dataset. Features include online activity metrics, demographic information, and historical academic performance, selected to capture diverse aspects influencing student success.

1.9 Normalization

Following feature extraction, normalization ensures standardized scales for features, preventing dominance by a single feature. Common methods like Min–Max scaling are applied to enhance the stability and convergence of predictive models [32].

1.10 Entropy and information gain

Entropy is a measure of impurity or disorder in a set of data. In the context of decision trees, entropy is used to determine the homogeneity or purity of a node. The formula for entropy (E) in the context of classification of n class problem is given by:

$$E = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Here, p_i represents the proportion of instances belonging to class i. The logarithm is typically base 2, and the negative sign is used to ensure a positive result. The entropy values range from 0 (completely pure node) to 1 (completely impure or mixed node). Lower entropy indicates a more homogeneous set of instances.

In the context of decision tree construction, when a node is split into child nodes, the information gain is calculated by comparing the entropy of the parent node with the weighted sum of the entropies of the child nodes. The formula for information gain (IG) is given by

$$IG = H(parent) - \sum_{j=1}^{k} \frac{N_j}{N} \cdot H(child_j)$$

where n is the number of classes, p_i is the proportion of instances belonging to class i at a particular node, $H(parent)$ is the entropy of the parent node, k is the number of child nodes after a split, N_j is the number of instances in the j-th child node, N is the total number of instances in the parent node and $H(child_j)$ is the entropy of the j-th child node.

1.11 Cluster analysis accuracy

Our study utilizes the k-means algorithm for cluster analysis. Evaluation involves grouping similar students, validating clusters, quantitative measures, interpretability, and practical significance. Results indicate successful grouping and meaningful patterns within clusters. K-means clustering is initiated by defining a cost function across a parameterized set of potential clusterings. The primary objective of the algorithm is to discover a clustering that minimizes this cost function. This clustering function is formulated as an optimization problem, wherein the objective function, denoted as$G$, operates on pairs of input (X, d) and a proposed clustering solution$C = (C_1, C_2 ..., C_k)$, yielding positive real numbers.

The task of the clustering algorithm is to identify, for a given input (X, d), a clustering C that minimizes $G((X, d), C)$. The objective function G is a key determinant in this process, and to achieve the goal, an appropriate search algorithm must be employed. K-means clustering, therefore, functions as a specific common approximation algorithm, rather than aiming for the exact solution to the minimization problem or the cost function.

Commonly, objective functions involve the parameter k, representing the number of clusters, and users often need to determine the most suitable value for k in practical applications. Several objective functions are utilized in clustering, and one of the most prevalent is the k-means objective function. This function quantifies the squared distance from each point in X to its cluster's centroid. In scenarios like digital communication tasks, where X members represent signals for transmission, the k-means objective function holds significance [33].

1.12 Chi-square parameter

The chi-square parameter $\left(\chi^2\right)$ in our study serves a specific purpose in the context of employing the CHAID algorithm for decision tree construction. It identifies optimal splits in the dataset based on the significance of relationships between variables, contributing to the formation of the decision tree structure. The chi-square parameter has purposes of optimal splits and node evaluation [34]. The optimal split is evaluated with the formula:

$$\chi^2 = \sum \frac{\left(O_i - E_i\right)^2}{E_i}$$

where O_i is the observed frequency in a particular category, E_i is the expected frequency in the same category. Node evaluation in the chi-square test is applied at each potential split to evaluate the homogeneity or heterogeneity of the target variable within the resulting nodes. This statistical criterion assists in identifying splits that lead to nodes with significantly different characteristics, improving the discriminatory power of the decision tree.

By employing the chi-square parameter in the CHAID algorithm, our research aims to create a decision tree that maximizes the predictive accuracy of the model by systematically identifying the most informative features and their interactions. This statistical rigor enhances the interpretability and reliability of the decision tree in predicting student success based on the specified criteria.

1.13 Comparison of academic success

Our research compares academic success before and during the COVID-19 pandemic across identified clusters. Parameters include success metrics, temporal comparison, and examination of success patterns, providing insights into variations in success rates.

2 Architecture of digital twin model

The digital twin model (UIS – University Information System) is based on a client/server web solution, where the end-user does not need to install a local mobile application. This application feature allows the proposed solution independent of a wide range of ICT devices and operating systems, allowing it to execute by any mobile or desktop device equipped with a standard web browser. Thus UIS was developed as a Web application, using various tools and technologies such as HTML5, CSS3, Jquery, C#, Ajax and Bootstrap, which executes with any standard web browser and serves as a communication interface between the user and the server via web services. The UIS interface we designed to be simple, attractive and customized for any user with quick access to all information.

The model presented in this paper use the MS-SQL database. The most important data like user names, status, NFC information and passwords in the database we store in the user tables. Password reset information is in the UserSequrity table. In the User table, redirection is made according to the required pages and the user's type. Thanks to the redirection, it is ensured that the students reach the student page, the professors to the instructor's page, and the student affairs workers to their page. Authorizations in the database are given according to the user type. In the StudentGrade table, the grades and achievements of the courses taken by the students are written. In the UserLang table, users' personal information in different languages is recorded. The DocRequest table contains the document information requested by the students.

The strategic challenge of one of the main problems of using new technologies in teaching activities is the ability to integrate three areas into a simple solution from the perspective of IT and users [35]:

- Information area, in which information should be handled and in what form it should be for quick and easy access of students in the teaching process.
- The area of functionality, i.e., what actions the student can take about this information and how these actions guarantee a consistent and efficient approach safely and satisfactorily by the requirements of the didactic activity.
- The teaching area is how to perform the instructional activity, rules, and regulations involved and sustained in each of the functional movements that students take to gain access to the information they need.

We resolve the integration of these three areas with the design of a web based model that gives a technological solution combining information, function, and requirements in the description and specification of all the actors involved. Each main item contains different groups under itself. This model, which consists of three main elements and sub-elements, exemplifies the use of digital twin technology in education as an open system. The content creation process, which is one of the main elements, includes more subgroups than reflected in the model; Individuals, professors and students also impact the content creation process. Kinsner states that disruptive technologies and ideas start in the educational process during the content creation phase [36]. The second main element in the model is described as the visible face of the digital twin, and it is explained that the knowledge and skills needed by the learner are provided in this element. In the model's last part, individualized personal needs distribution is made. Evaluations can be made in this element in line with the context, time, and needs.

The user enters the University Information System web page. The login section reads the NFC ID card from the phone or to the NFC card reader connected to the computer. This information becomes authentication in Active Directory. If the information is correct, the Single-Sign-On authentication scheme allows access to the application (UIS) and Office 365. As seen in Figs. 1, 2 and 3, the interface is organized with two-step verification and provides access to the main elements of data according to user attributes.

UIS is a web-based application designed to support the educational process at International Vision University. The operation and management (knowledge base) subgroup of our symbiotic model contains the following entities or elements:

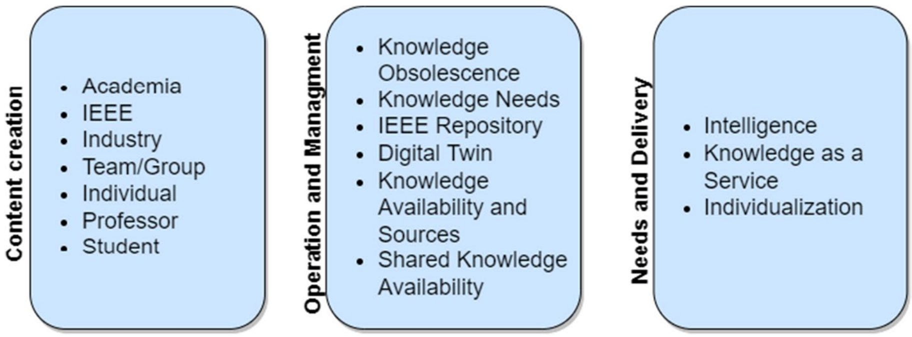

Fig. 1 A Digital Twin Model for Symbiotic Education

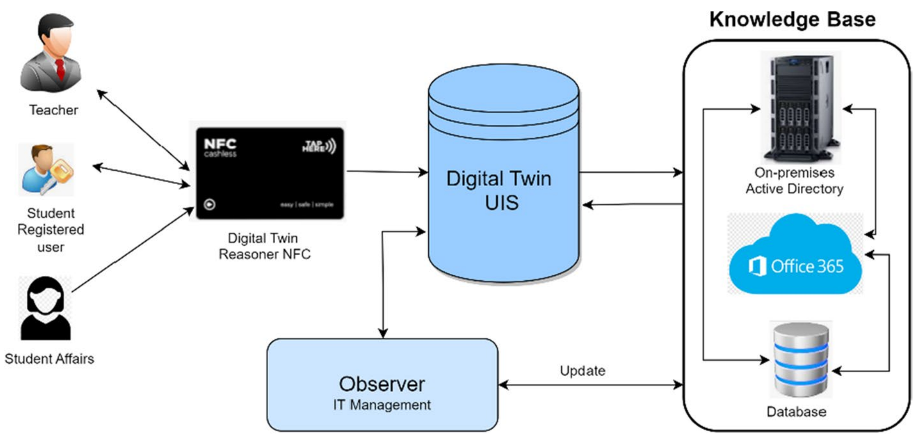

Fig. 2 Digital Twin Model Architecture

Fig. 3 Our Model as Mobile and Web Application

2.1 Reports

Student affairs referents and instructors are allowed to enter this entity. Students affairs referents have the authority to receive student documents, student lists, filtered student lists, transcripts and diplomas (printed versions). At the same time, the instructors have permission to see the student lists, class participation, and grades.

2.2 Student information

Only student affairs referents are allowed to access this entity. Here, Student Affairs has the authority to enter and change all information about students (ID number, faculty, department, year, personal information- can reference the database).

2.3 Student groups

Only student affairs referents have permission to enter into this entity. Student groups are formed here by considering features such as faculty/academic year/department. These groups are assigned to the courses stipulated by the department's curriculum in the academic year.

2.4 Courses groups

Student affairs, instructors, and students can enter this entity. This entity has access to professors and students. Student affairs only have the authority to create groups here and export the created groups to Teams in.csv format. The group owners are the professors who are lecturers in the particular group.

2.5 Lectures participation

Only instructors can enter here; they have permission to upload course participation points.

2.6 Projects

Only instructors can access this entity. In this entity, instructors upload the project points.

2.7 Exams

The instructor writes midterm, final, and make-up exam scores here. Students are only authorized to see the scores entered by the instructor.

2.8 Other activities

At this entity, the instructors write scores for the student's actions that correlate with the subject's nature (participation in seminary, panel discussion, scientific visit, etc.). Students are only authorized to see the scores entered by the instructor.

Lecture notes (document management): Instructors are authorized to upload course materials (lecture notes, presentation files, video) in this unit. Students can download the course materials uploaded by the instructors.

2.9 Grades

The exam scores entered by the instructor here are evaluated and it is decided whether the student has passed/failed the exam.

3 Testing the impact of uis on the educational process

The *k*-means algorithm clusters data by identifying relationships and patterns within the training dataset. In this study, we applied the *k*-means algorithm to group students with similar online activity and success levels, identifying four distinct clusters. Analyzing all 13 inputs, we found that Faculty and Grade are the two factors with the highest predictor importance in our dataset, see Figs. 4 and 5.

Among the obtained clusters, the smallest one comprises 113 instances, accounting for approximately 10.7% of the total dataset, as illustrated in Fig. 4. This smaller cluster suggests the presence of a distinct subgroup within the overall population. On the other hand, the largest cluster, consisting of 338 instances, represents a significant 32.1% of the dataset, indicating a prevalent pattern or characteristic within this substantial portion of the population.

The ratio of the largest cluster to the smallest cluster is calculated at 2.76, revealing a noticeable difference in size between these clusters. However, this ratio is smaller than in other clustering scenarios, suggesting a more balanced distribution. This implies a more

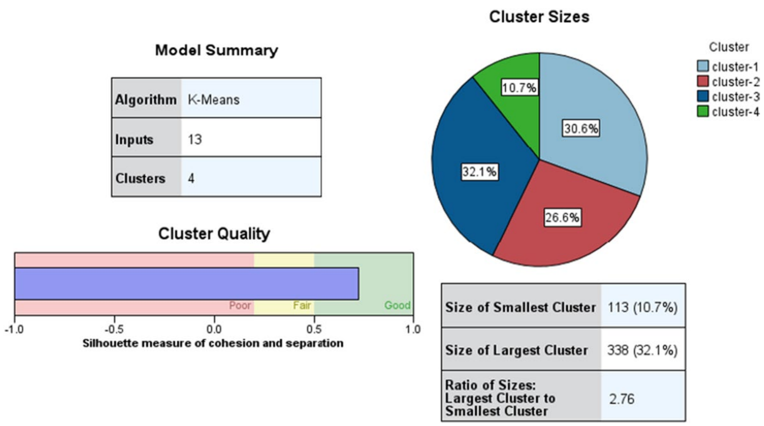

Fig. 4 Clustering of the 13 inputs data

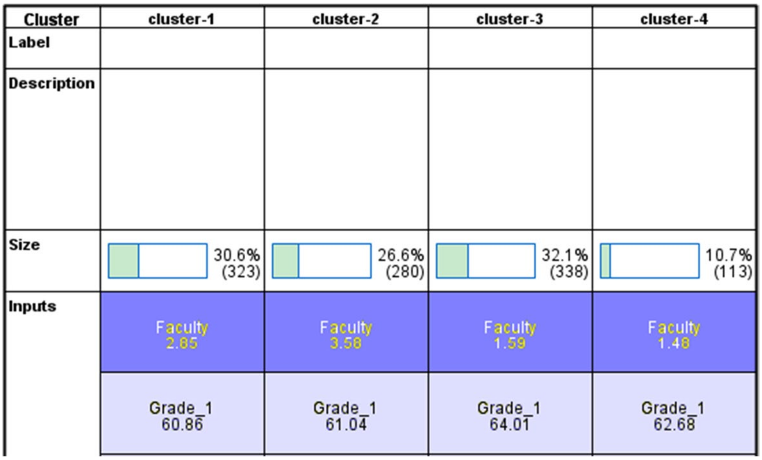

Fig. 5 Clusters for Faculty factor

even distribution of instances across the clusters than scenarios with higher ratios, indicating a balanced representation of distinct patterns or behaviors within the dataset.

While the difference in cluster sizes is notable, it is less extreme than in some cases, pointing towards varying patterns or characteristics among different population subsets. Overall, the clustering analysis provides insights. The primary goal of cluster analysis is to group students with similar characteristics, particularly in online activity and success levels. Employing the *k*-means algorithm, which identifies clusters by minimizing the within-cluster sum of squares, we successfully grouped students into four clusters based on online activity and success levels.

A thorough analysis of the characteristics within each cluster was conducted to validate the accuracy of the obtained clusters. This involved examining the distribution of online activity metrics and success levels for each identified group. Quantitative measures, such as silhouette analysis and within-cluster sum of squares, were utilized to assess the compactness and separation of the clusters. Silhouette analysis, in particular, provided insights into the appropriateness of the clustering solution by measuring how similar an object is to its own cluster compared to other clusters.

Beyond quantitative measures, consideration was given to the interpretability and practical significance of the obtained clusters. Meaningful and distinct clusters with clear characteristics enhance the useful utility of the analysis. The analysis revealed distinct patterns within each cluster, providing valuable insights into the relationships between online engagement and academic performance.

Quantitative measures indicated a high degree of compactness within and separation between clusters, suggesting that the k-means algorithm effectively captured underlying patterns in the data. The results demonstrated that students within each cluster shared similar traits, validating the accuracy of the cluster analysis in identifying meaningful groupings.

Moving forward, we comprehensively compared academic success in exams before the pandemic (face-to-face methods) and during the COVID-19 pandemic (online methods) across the four identified clusters. Success rates in exams were compared, considering defined grading criteria in the context of the law on higher education in North Macedonia.

The temporal comparison allowed us to discern any variations in success patterns across the identified clusters during face-to-face and online learning environments. Statistical tests, such as chi-square tests, were applied to determine the significance of observed differences in success rates between the two periods, enhancing the reliability of our findings.

The temporal comparison allowed us to discern any variations in success patterns across the identified clusters during face-to-face and online learning environments. Statistical tests, such as chi-square tests, were applied to determine the significance of observed differences in success rates between the two periods, enhancing the reliability of our findings.

Our analysis considered the implications of the findings, discussing factors that may contribute to observed differences or similarities in success rates. This included potential influences of online learning modalities, changes in student engagement, and other contextual elements. The entire research approach ensures a robust and accurate understanding of the relationships between online activity, success levels, and student groupings, both in the clustering analysis and the subsequent comparison of success rates. Into the diverse patterns and behaviors present within the dataset.

To enrich the comprehensiveness of our exploration, the presented analysis integrates entropy and CHAID tree methods to unveil patterns and relationships within a dataset, focusing specifically on key factors such as gender, grades, and academic years. Entropy serves as a metric to gauge the level of uncertainty associated with various dependencies within the data. Simultaneously, the CHAID tree method provides a hierarchical breakdown of the dataset, revealing intricate connections between different characteristics.

The entropy results from our analysis, as shown in Table 1, showcase low levels of uncertainty in categories related to gender, academic years, and grades. This suggests a certain level of predictability in these aspects, offering valuable insights into the stability and consistency of the dataset across these key variables.

Building upon this, the CHAID tree analysis further dissects the dataset, providing a detailed view of how gender and grade characteristics interplay across different academic years and periods. The hierarchical nature of the CHAID tree reveals nuanced relationships, contributing to a more thorough understanding of the dynamics within the dataset.

The ’Gender’ decision tree was constructed based on the training subset ($n = 150\ 472$) and tested and validated with the validating subset ($n = 100\ 988$). The resulting tree contained twenteen nodes, of which eight were leaf nodes. The fit statistics indicated that 96.4% of all the training and validating records were correctly classified, resulting in a classification rate of 0.38 (see Fig. 6).

The nodes in the decision tree diagram display the frequencies of the target variable as a percentage, counts of identified predictors, along with the total number of cases analyzed at each node. Starting with 150 472 records in the training subset, the decision tree algorithm identified the item ’Grade’ as the most important input for prediction, followed by the ’Pre

Table 1 Entropy Results for Gender -Academic Year-Grade categories

Value		
Uncertainty Coefficient	Symmetric	.006
	Grade Dependent	.005
	Gender Dependent	.010
Uncertainty Coefficient	Symmetric	.000
	Year Dependent	.000
	Gender Dependent	.000

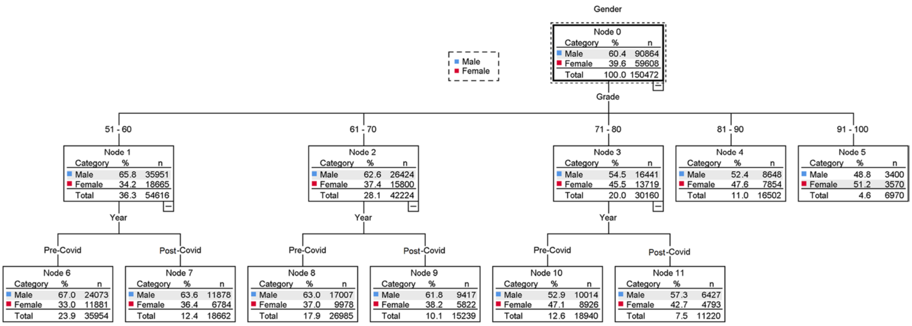

Fig. 6 Decision tree for the prediction success by attribute Gender

Covid—Post-Covid' item. The root node was initially partitioned into five child nodes based on the grades obtained from exams. Nodes one and three (children of root/Gender node) became parent nodes with the decision tree algorithm, identified as significant inputs to success in 'Pre-Covid—Post-Covid' items for the 'Gender' attribute of students. These child nodes were identified as leaf nodes, indicating that there is not a significant difference in success by gender attribute, leading us to conclude that online courses are effective for both genders.

The entropy results in Table 2 highlight the uncertainty and dependencies within the dataset, focusing on symmetry and factors such as academic years, grades, and faculty. The Uncertainty Coefficient, a symmetric measure, is 0.010, indicating a general level of uncertainty. When examining academic years, the uncertainty remains consistent at 0.010. However, concerning faculty dependence, the uncertainty slightly increases to 0.011, suggesting a nuanced and potentially more complex relationship with the target variable.

In the symmetric analysis, the Uncertainty Coefficient drops to 0.009, indicating a lower overall level of uncertainty compared to the symmetric approach. Grade Dependence shows a decrease to 0.006, suggesting a relatively more predictable relationship associated with grades. Conversely, Faculty Dependence sees an increase to 0.015, indicating a potentially more intricate and less predictable connection with the target variable.

These entropy results contribute to a comprehensive understanding of the dataset, emphasizing varying degrees of uncertainty and dependencies associated with different factors, including academic years, grades, and faculty. The analysis also suggests a potential avenue for further exploration with the decision tree.

Transitioning to the 'Faculty' decision tree, constructed based on the training subset ($n = 150\ 472$) and tested with the validating subset ($n = 100\ 988$), the resulting tree has

Table 2 Entropy Results for Faculty -Academic Year-Grade categories

Value		
Uncertainty Coefficient	Symmetric	.010
	Year Dependent	.010
	Faculty Dependent	.011
Uncertainty Coefficient	Symmetric	.009
	Grade Dependent	.006
	Faculty Dependent	.015

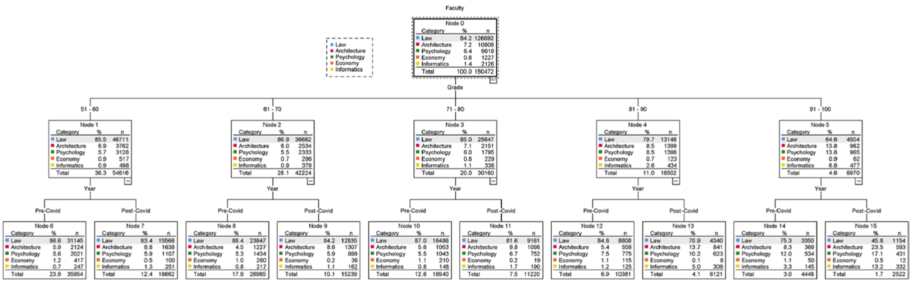

Fig. 7 Decision tree for the prediction success by attribute Faculty

fifteen nodes, with nine being leaf nodes. Fit statistics indicate 98% accuracy in correctly classifying all training and validating records, resulting in a classification rate of 0.02 (see Fig. 7).

The decision tree diagram illustrates target variable frequencies, identified predictors, and the total number of cases analyzed at each node. Starting with 150,472 records in the training subset, the decision tree identifies 'Grade' as the most important predictor, followed by the 'Pre Covid—Post-Covid' item. The root node initially splits into five child nodes based on grades, each containing all grades for five faculties. This high probability derived from students' grades suggests that the UIS application significantly impacts the educational process at International Vision University. The success during the pandemic, evident in the decision tree leaves, underscores the positive influence of this application on academic outcomes.

4 Discussion

In our study, it's important to examine the impact of incorporating conceptual digital twin models in education, as suggested in prior research [35]. These models have significantly enhanced the learning experience, especially in open and distance learning. The integration of symbiotic digital twins and scattered information in instructional design provides valuable tools for integrating learner information, making predictions, and delivering accurate evaluations and rapid feedback [37].

The literature review emphasizes the potential of digital twin technology in open and distance education systems, enabling quick responses to immediate changes and proactive addressing of future challenges. Our study aligns with this perspective, highlighting digital twin technology as a powerful system applicable not only in open and distance learning but also in various domains where its capabilities are needed.

Our findings contribute to the discourse on the application of digital twin models in education. The evaluation data, especially in the decision trees, show no significant differences in student success across all five faculties in both face-to-face and online teaching scenarios. This suggests that the digital twin model, represented by the UIS application, plays a crucial role in realizing the educational process, especially during pandemics.

The positive impact of the digital twin model is further emphasized by the wide acceptance among our students. The individualized and adaptive learning approach facilitated by our digital twin model introduces a novel way of dynamically and effectively assessing students' knowledge levels. The personalized nature of the model proves particularly beneficial for international students, offering substantial support across various aspects of online education during the pandemic.

Furthermore, the model adheres to general design principles of cybersecurity and encompasses elements specific to the implementation of distance learning. The ability to personalize our model not only enhances the learning experience but also serves as significant support for students, especially those studying remotely. The successful realization of the educational process during the pandemic underscores the positive influence of the digital twin model at International Vision University.

While our study provides valuable insights and evidence, it has several limitations. The findings are based on data from a specific institution, limiting generalizability to other educational settings. The study relies on the dataset's availability and quality, potentially impacting the conclusions drawn. It primarily focuses on a specific time frame during the COVID-19 pandemic, and external factors could influence the results over time. The positive impact of the digital twin model depends on the technology and context of International Vision University, affecting its applicability in other institutions. The study assumes a positive reception of the digital twin model by students, without addressing variations in individual preferences and adaptability. It highlights benefits without delving into specific digital twin implementations, and the discussion overlooks external factors and ethical considerations. Additionally, the study predominantly relies on quantitative data, lacking qualitative insights for a more comprehensive understanding of students' experiences. Incorporating qualitative research methods could offer a more holistic view.

5 Conclusion

In conclusion, our study illuminates the significant potential of digital twin technology as a valuable tool for enriching the learning process. The findings underscore the benefits of integrating this technology into instructional design processes and learning environments, empowering educators to gather accurate assessments through the comprehensive collection of diverse learner profiles. As a simulation-based technology, digital twin offers a spectrum of enriching application opportunities, enabling educators to design teaching environments that harness its relevance effectively.

Furthermore, our study highlights the imperative for continued exploration and research in this evolving field. We encourage educators and researchers to embark on experimental or design-based studies investigating the application of digital twin technology in education, thereby charting new directions for future endeavors. We specifically propose exploring the conceptual link between digital twins and simulation theory, acknowledging the technology's roots in this theory. Addressing the gap in research on the optimal integration of digital twin technology with different technologies in the learning process holds the potential for a more comprehensive understanding of its synergies with diverse educational methodologies.

Given the unprecedented challenges posed by the Covid-19 pandemic, we advocate for studies that probe the utility of digital twin technology across various teaching modalities, encompassing face-to-face, distance, and hybrid education. The post-pandemic era necessitates a thorough examination of how digital twin technology can be effectively employed in the teaching process, requiring in-depth analyses of its functionalities along with assessments of advantages and disadvantages. Our research, incorporating the addition of chat rooms in UIS, serves as a stepping stone for future investigations. We plan to evaluate this functionality, utilizing neural network analysis and k-means clustering to discern its advantages and disadvantages.

In essence, our study not only accentuates the current benefits of digital twin technology but also serves as a catalyst for ongoing exploration and research. We envision a future where the integration of this technology in education evolves to become more refined, versatile, and tailored to the diverse needs of learners and educators. The pursuit of these future directions promises to unlock new dimensions in the application of digital twin technology, shaping the landscape of education in innovative and meaningful ways.

Author contributions Aybeyan Selim: Conceptualization, Investigation, Methodology, and Writing – original draft; Ilker Ali, Data curation, Resources, and Visualization; Muzafer Saracevic and Blagoj Ristevski: Supervision.

Data availability Data sets generated during the current study are available from the corresponding author on reasonable request. The data are available from the Academic Information System of International Vision University database from North Macedonia; restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Declarations

Ethics committee approval This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of International Vision University with Ref. 2002–97/03. Confidentiality of the participants was maintained throughout the study, and all data collected were stored securely and analyzed anonymously.

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Kumar D, Sood SK, Rawat KS (2023) Early health prediction framework using XGBoost ensemble algorithm in intelligent environment. Artif Intell Rev 56(Suppl 1):1591–1615. https://doi.org/10.1007/s10462-023-10565-6
2. Hermann M, Pentek T, Otto B (2016) Design Principles for Industrie 4.0 Scenarios. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). Koloa, HI, USA, pp 3928–3937. https://doi.org/10.1109/HICSS.2016.488
3. Radouan Ait Mouha RA (2021) Internet of things (IoT). Journal of Data Analysis and Information Processing 9:77–101
4. Luigi A, Antonio I, Giacomo M (2010) "The Internet of Things: A survey. Comp Net 54(15):2787–2805
5. S. Kumar Sood, K. Singh Rawat and D. Kumar, "Emerging Trends of ICT in Airborne Disease Prevention," ACM Transactions on Internet Technology, vol. 22, no. 4, pp. 1 - 18, 2023.
6. Grieves M (2014) Digital Twin: manufacturing excellence through virtual factory replication. LLC
7. Bréchet L, Mange R, Herbelin B, Theillaud Q, Gauthier B, Serino A, Blanke O (2019) First-person view of one's body in immersive virtual reality: influence on episodic memory. PloS One 14:e0197763
8. Kumar D, Sood SK, Rawat KS (2023) Empowering elderly care with intelligent IoT-Driven smart toilets for home-based infectious health monitoring. Artif Intell Med 144:102666. https://doi.org/10.1016/j.artmed.2023.102666
9. Domb M (2019) Smart home systems based on internet of things, IoT and Smart Home Automation. IntechOpen. https://doi.org/10.5772/intechopen.84894
10. Noueihed H, Harb H, Tekli J (2022) Knowledge-based virtual outdoor weather event simulator using unity 3D. J Supercomput 78:10620–10655. https://doi.org/10.1007/s11227-021-04212-6
11. Bhagavathi R, Kufoalor DKM, Hasan A (2023) Digital twin-driven fault diagnosis for autonomoussurface vehicles. IEEE Access 11:41096–41104. https://doi.org/10.1109/ACCESS.2023.3268711
12. Chen D, Wang D, Zhu Y, Han Z (2021) Digital twin for federated analytics using a bayesian approach. IEEE Internet of Things J 8(22):16301–16312. https://doi.org/10.1109/JIOT.2021.3098692

13. S. Hung, D. Yen and H. Wang, "Applying Data Mining to Telecom Churn," Expert Systems With Applications, vol. 2005, pp. 1–10.
14. Wang Z (2020) Digital twin technology. In: Industry 4.0 - Impact on Intelligent Logistics and Manufacturing. IntechOpen, pp 95–114. https://doi.org/10.5772/intechopen.80974
15. Sekione RJ, Laurence TY, Jong HP (2023) Digital twin-assisted resource allocation framework based on edge collaboration for vehicular edge computing. Futur Gener Comput Syst 150:243–254
16. Ebner M, Maierhuber M (2013) Near field communication - which potentials does NFC bring for teaching and learning materials? Int J Interact Mob Technol 7(4):9–14
17. Coskun V, Ozdenizci B, Ok K (2013) A survey on near field communication (NFC) technology. Wireless Pers Commun 71:2259–2294. https://doi.org/10.1007/s11277-012-0935-5
18. Shen CW, Wu YCJ, Lee TC (2014) Developing a NFC-equipped smart classroom: effects on attitudes toward computer science. Comput Hum Behav 30:731–738. https://doi.org/10.1016/j.chb.2013.09.002
19. Johannes V (2020) NDE Perception and Emerging Reality: NDE 4.0 Value Extraction. Mat Eval. 78(7):835–851
20. Selimi A, Useini A (2019) Yenilikçi Eğitim ile Dijital Yetkinlik ve Girişimcilik Becerilerinin Geliştirilmesi - Kuzey Makedonya Örneği," In: ICEB'19 - International Congress of Economics and Business, pp 204–213
21. James K, Humphrey G, Vilis T, Corrie B, Baddour R, Goodale M (2002) "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. Behav Res Methods Instrum Comput 34:383–390
22. Horne M, Thompson E (2008) The role of virtual reality in built environment education. J Educ Built Environ 3:5–24
23. Lin C-C, Tsai C-C (2009) The relationships between students' conceptions of learning engineering and their preferences for classroom and laboratory learning environments. J Eng Educ 98:193
24. Yeary M, Yu T, Palmer R, Biggerstaff M, Fink L, Ahern C (2006) A hands-on, interdisciplinary laboratory program and educational model to strengthen a radar curriculum for broad distribution. In: Proc. ASEE Annu Conf Expo, pp 11.52.1–11.52.20. https://doi.org/10.18260/1-2--129
25. Schuster K, Groß K, Vossen R, Richert A, Jeschke S (2016) Preparing for industry 4.0 – collaborative virtual learning environments in engineering education. Available: https://link.springer.com/chapter/10.1007/978-3-319-46916-4_36. Accessed 9 1 2024
26. Toto R, Hien Nguyen (2009) Flipping the work design in an industrial engineering course. In: 2009 39th IEEE Frontiers in Education Conference. San Antonio, TX, USA, pp 1–4. https://doi.org/10.1109/FIE.2009.5350529
27. Selimi A, Saračević M, Useini A (2020) Impact of using digital tools in high school mathematics: a case study in north macedonia. Univers J Educ Res 8(8):3615–3624
28. David J, Lobov A, Lanz M (2018) Leveraging digital twins for assisted learning of flexible manufacturing systems. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN). Porto, Portugal, pp 529–535. https://doi.org/10.1109/INDIN.2018.8472083
29. V. Vikhman and M. Romm, ""Digital twins" in Education: Prospects and Reality. Vysshee obrazovanie v Rossii = Higher Education in Russia.," vol. 30, no. 2, pp. 22–32, 2021.
30. I. Kolesnikova, "Post-Pedagogical Syndrome of the Era of Digital Modernism," Vysshee obrazovanie v Rossii = Higher Education in Russia. 2019 28 (89);67–82,
31. Medina IJ, Gómez-Luque MA, Peña Amaro J, Luque Ruiz I, Gómez-Nieto MA (2019) HistoNFC: an innovative tool for the practical teaching of histology using NFC technology. Wirel Commun Mob Comput (2019):1–16. https://doi.org/10.1155/2019/1291364
32. Raschka S (2015) Python machine learning. Packt Publishing Ltd, Birmingham, UK
33. Shalev-Shwartz Sh, Ben-David Sh (2014) Understanding machine learning: From theory to algorithms. Cambridge University Press, Cambridge, United Kingdom
34. Breiman L, Friedman J, Olshen RA, Stone CJ (1984) Classification and regression trees, 1st edn. Chapman and Hall/CRC
35. Ferrara G, Ferrari L, Vichi G, Varrochi A (2011) A methodology for engine performance optimization. SAE Technical Paper, pp 1–12
36. Dambrot SM (2019) Symbiotic autonomous systems, digital twins and artificial intelligence: emergence and evolution. Mondo Digitale, pp 1–14
37. Kinsner W (2019) Towards evolving symbiotic education based on digital twins. Mondo Digitale 2:1–14