Applications of machine learning in operational aspects of academia: a review

Muhammad Nadeem, Wael Farag, Zekeriya Uykan, Magdy Helal
College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait

ABSTRACT
Educational institutions, propelled by digital transformation and sophisticated machine learning (ML) algorithms, amass plentiful data, facilitating the execution of complicated decision-making tasks previously inconceivable. ML’s pervasive influence extends beyond pedagogy and research, profoundly altering the fabric of academia and reshaping university functionalities. Its deployment in university administration enhances efficacy, efficiency, and operational streamlining across diverse levels. This article conducts a comprehensive review of extant knowledge pertaining to the diverse applications of ML in non-teaching domains within academic settings, delineating avenues for future research. The recognized findings furnish a robust foundation for the further exploration and refinement of ML applications, particularly within the administrative and operational realms of academia. A consequential outcome of this transformative integration is the mitigation of teachers’ administrative burdens. In practical terms, this liberation affords educators the opportunity to redirect their time and energy towards their primary responsibilities of educating and fostering the intellectual development of their students.

Keywords:
Administration operations
Digital transformation
Higher educational institutes
Machine learning
Operational aspects
University administration

This is an open access article under the CC BY-SA license.

Corresponding Author:
Muhammad Nadeem
College of Engineering and Technology, American University of the Middle East
Egaila, 54200, Kuwait
Email: muhammad.nadeem@aum.edu.kw

1. INTRODUCTION
The use of machine learning (ML) has become necessary in almost all fields of life. ML provides useful insights into the current education system that have not been seen before. The reason is the ML’s ability to convert data into intelligent actions. All educational institutions gather significant amounts of data, at all levels, usable for assessments and decision-making purposes. It is, therefore, natural to utilize ML for developing and obtaining more useful insights out of the available data, for the benefit of the students and the academicians as well as the educational management systems [1]. The world bank reckoned that the investments in artificial intelligence (AI) applications in education have crossed the figure of USD 1 billion from 2008 to 2019 [2], [3]. The significance of AI usability in education has been emphasized by UNESCO and the OECD, for sustainable development in society [4]. The market size of ML services is expected to grow beyond USD 300 billion by 2030 as shown in Figure 1. Its applications have exploded in various fields of study such as healthcare [5] and education [6], [7].

Machine learning is a branch of AI that aims at developing models and algorithms that allow computers to learn from given datasets and become better at a given task without having to be explicitly programmed. It enables computers to draw conclusions or take actions based on what they have learned from past experiences or historical data. ML is a broad science with many subfields, each of which focuses on
Machine learning techniques have become widely used in a variety of educational contexts because of their propensity to analyze huge datasets and spot patterns. This includes individualized instruction, flexible testing, and sophisticated tutoring tools, all of which improve the teaching and learning process. Research on implementing ML in the education sector is accelerating. It presents and will present new perspectives, and help in developing enhanced, effective tools with numerous benefits. In recent years, ML, and AI in general, have been widely implemented in the education systems; in K-12 [9], higher education [10], and intelligent tutoring systems. But ML has also been applied to the non-teaching aspects of academia which includes activities and procedures that support the educational ecosystem but are not directly connected to pedagogical activities. These include administrative duties, student services, and research projects aimed at raising the standard of education.

An academic job typically includes a variety of components, such as teaching, research, service, administration, professional development, and outreach. Teaching is the main component of an academic job which involves developing and delivering lectures, leading discussions, designing, and grading assignments and exams, and providing mentorship and support to students. It is also required to make service contributions to the department, institution, or broader academic community. This can include serving on committees, participating in academic events, reviewing papers and grant proposals, and other activities that support the academic mission, in addition to administrative responsibilities, such as serving as department chairs, program directors, or deans. These roles require skills in management, leadership, and strategic planning. Although these administrative tasks are an essential part of academic life, they can have several negative consequences for lecturers. These tasks are time-consuming which may result in reduced teaching time, lowered research output, diluted teacher's expertise, and burnout. Integrating ML solutions into administrative activities may solve this problem and reduce the workload of lecturers thus allowing them to concentrate more on research, teaching, and engaging with students. ML can be applied to all aspects of academia, but the scope of this paper is limited to the service and administration aspect while the remaining components are not kept out of the scope of this paper to make it concise.

Increasing numbers of educational institutions have shown interest in applying ML for making important decisions regarding admission, curriculum design, student registration, and retention. This tendency towards integrating ML in the operations can be leveraged in both the teaching and administrative aspects of the education sector. For instance, a learning management system (LMS) is an ideal candidate for applying these technologies and improving the learning experience. There are few studies reported [11], [12] where ML was integrated with LMS for academic support and student performance management. Applying such predictive technologies to the abundant data provided by this system can dramatically increase the efficacy of the system. This implies that individuals running the institution and making executive choices will be able to do so with more knowledge.
The exploration of ML potential benefits in non-teaching areas of education has yielded valuable insights, as indicated by existing research. However, the dynamic and evolving nature of this field suggests the need for further investigations in order to achieve a thorough understanding of these advantages. Delving deeper into the intricacies and nuances, additional research will help uncover the full spectrum of benefits while also shedding light on any potential drawbacks associated with the implementation of ML in diverse educational contexts. This extended inquiry is crucial for informing decision-makers, educators, and policymakers on how to harness the power of ML optimally and address any challenges that may arise.

This review article focuses on the administrative and support functions in universities, exploring the reported application of the ML approaches in those functions. All authors have academic backgrounds and based on their experience and literature review, the most significant areas of applications identified were predicting and managing activities in the areas of student enrollment, learning disabilities, attendance tracking, selection of the program of study/university, class timetabling, loans and financial aid allocation, and graduation project planning, and in the curriculum, development are explored in the following sections.

2. METHOD

This study constitutes a literature review, specifically a narrative review, elucidating the existing knowledge concerning the implementation of ML in non-teaching facets of university environments. The research methodology adopted for this endeavor encompassed a meticulous examination of extant literature and case studies relevant to ML roles in non-teaching dimensions of education. Primary data collection involved searching and inspecting real-world instances of ML applications within educational institutions.

Notably, paper inclusion or exclusion criteria were not predicated on publication sources, but rather, a language-based criterion was imposed, restricting inclusion to papers published in the English language. A systematic search strategy was employed, leveraging ML-related keywords to retrieve pertinent literature on the targeted subject. Electronic databases, including but not limited to Scopus, Web of Sciences, Google Scholars, IEEE Xplorer, association for computing machinery (ACM), and SpringerLink, were thoroughly explored to compile a comprehensive selection of relevant materials.

To ensure the quality and breadth of the reviewed articles, various formats such as journal articles, academic conference papers, workshop materials, book chapters, and English-language lecture notes or talks were considered. A screening was made with a careful examination of titles and abstracts to identify potentially germane sources, excluding those unrelated and contrasting with the research question. In a subsequent screening phase, papers lacking information on methodology, key findings, and/or experiments were excluded. The final stage of the review involved a comprehensive scrutiny of the full texts of selected sources, assessing their methodology, key findings, and relevance to the research objectives. The collated material was methodically categorized and synthesized to offer a comprehensive overview of the ML applications within the non-teaching realms of academia.

3. RESULTS AND DISCUSSION

In this section, we present the key findings derived from our study investigating the application of ML and non-teaching settings of higher education institutions (HEI). The data collected through (describe methods briefly), reveal significant insights into the ML application for predicting and managing activities in the areas of student enrollment, learning disabilities, attendance tracking, selection of the program of study/university, class timetabling, loans and financial aid allocation, and graduation project planning, and in the curriculum development. We begin by providing a summary of the key results and their relevance to our research goals. Subsequently, we discuss the challenges faced and future trends. While interpreting the results, we also acknowledge certain limitations inherent in our study. This section sets the foundation for a comprehensive discussion of the study's outcomes, shedding light on their implications and contributing to the body of knowledge.

3.1. Applications of machine learning in administration and operation

Several areas of application of ML in the administrative or non-teaching aspects of academia were identified and are presented in this section. These applications demonstrate the transformative potential of ML in revolutionizing administrative processes within academic institutions, leading to increased efficiency, improved decision-making, and enhanced overall performance. Some of the applications include predicting enrollment, student attendance, university selection, timetabling and scheduling, student loans, and final project allocation. These advancements underscore the significant impact that ML can have on the efficiency and effectiveness of academic administration.
3.1.1. Predicting enrollments

Accurately predicting academic enrollment in HEI is of paramount importance. It is the basic input for preparing the necessary resources for the new students. Accurate enrollment prediction brings efficiency to the entire process since unnecessary administrative workload can be reduced and every university wishes to grant admission to eligible applicants who are ready to enroll in the institution that extends the offer. On the one hand, it is conceivable that difficulties will arise in the future if students who are not qualified or motivated are admitted. Failure to meet enrollment criteria puts universities' financial stability in jeopardy, which is a particularly critical issue for private colleges. The institution's annual budget is significantly dependent on student enrollment. In many cases, courses are either over-enrolled or under-enrolled [13]. Both situations can potentially create issues for the students and institutions. Traditionally, academic institutes confirm the students’ enrollments by their payment status. However, this puts a lot of pressure on the administrative staff and leaves little to no time for institutes to be fully prepared for the semester/year. Another solution is to make a rough estimate of enrollments using the previous years’ data. However, that might not present accurate information [14], [15]. Therefore, with the advances of ML and the availability of data, early and accurate predictions can potentially be forecasted using ML by analyzing each candidate individually and forecasting their likelihood of accepting admission. The ML model functions by considering parameters about the applicant, such as their high school GPA and test scores, and by looking at patterns among applicants who are like them from past years. Additionally, ML models can foretell which students would flourish at a specific university. Many earlier approaches to choosing candidates used formulaic calculations to ascertain a candidate’s likelihood of success which can be enhanced with ML to make the final forecasts more accurate.

To predict students’ enrollment, it is important to have access to student admissions data. Most institutions have abundant data, but infrastructure restrictions and untrained staff make it difficult to analyze the data. As a result, institutions require the assistance of private consulting companies. Therefore, proper processing facilities should be made available in addition to data. Institutions may be “rich” in the data context but “poor” in data processing, mostly due to infrastructure restrictions and untrained staff [16]. Consequently, institutions require the assistance of private consulting companies, which do not share how their results are evaluated.

Several studies have investigated the use of ML for students’ enrollment. Aulck et al. [17] used a genetic algorithm to demonstrate how the existing students’ data can be used to increase enrollment and optimize enrollment management. Firstly, ML classifiers are used to predict first-year enrollments. Based on this prediction, a genetic algorithm was used to optimize the results. Canada et al. [18] took a different approach and utilized logistic regression (LR) to identify the factors contributing to the students’ decisions to enroll to a specific program. Various classification algorithms were also benchmarked, but LR proved to provide the best results. James and Weese [19] investigated the future enrollment patterns using a 5-year student performance record. A neural network-based time forecasting series was found to be the best model. Slim et al. [20] used semi-supervised probability techniques, support vector machines (SVM), and LR to predict enrollments. The enrollment of each student is predicted using LR and SVMs, while the enrollment is predicted using semi-supervised probability techniques. The University of New Mexico (UNM) provided the actual data for 54,692 students used in this study between 2009 and 2016. According to the findings, students who get state grants enroll at UNM at a higher rate. Additionally, by utilizing just a limited number of parameters linked to college and student characteristics, enrollment may be predicted using the suggested models with sufficient accuracy. Similar research utilizing Adaboost and Random Forest (RF) [21]. Yang et al. [14] suggest combining the Whale Optimization with Whale Optimization Algorithm (WOA) and Support Vector Regression (WOASVR) to forecast and analyze student enrollment and teacher statistics in Taiwan. Yang et al. [15] collected four years of first-year enrollment data and the first three years’ data was used for training purposes whereas the fourth year’s data was used for testing. They applied RF and neural networks (NN) to train the model. The meaningful enrollment information can be obtained with relatively good accuracy. Sghir et al. [22] compared the accuracy of three ML algorithms; decision tree (DT), RF, and SVM for predicting the new student enrollments. They found that RF is more suited to solving such problems. Golden et al. [23] examined and contrasted various ML approaches for forecasting university admissions.

Predicting enrollment using ML can be complex and challenging because of a variety of factors, such as changes in demographics, the economy, and global events. Previous enrollment data can be incomplete, inaccurate, or inconsistent, which can affect the accuracy of the predictions made by ML algorithms. Enrollment data can be complex and multidimensional, and include variables such as student demographics, application data, historical enrollment trends, and external factors. ML algorithms may struggle to identify relevant variables or to account for interactions between variables. Historical enrollment data may not be available for newer programs or institutions, which can limit the effectiveness of ML models. Enrollment prediction with ML raises ethical concerns regarding privacy, fairness, and bias. ML models may perpetuate or amplify existing biases if they are not appropriately designed or trained.
3.1.2. Student class attendance

Many higher education institutes require keeping track of students’ attendance due to a strong correlation between attendance and performance [24]. However, keeping track of students’ attendance is difficult, especially for large class enrollment. The manual attendance methods are tedious, time-consuming, and prone to errors and its inefficacy was evident during the COVID-19 era where academia faced challenges in taking attendance [25]. Sometimes, students fill in for absent friends by acting on their behalf. ML can be used to track and analyze class attendance data, which can provide valuable insights for instructors and administrators. Once attendance data is collected, ML can be used for analysis and identification of present patterns and trends that may indicate potential issues or opportunities for improvement. For example, it can help identify students who are frequently absent and may need additional support or interventions. ML can be applied to forecast future attendance patterns based on historical data [26], which can help instructors and administrators plan and allocate resources more effectively.

Most of the work related to the attendance system is published in conferences and most of the researchers are in India. Researchers have used approaches based on biometrics and non-biometrics. The biometric-based approaches used for attendance tracking include fingerprint [27], iris [28], GPS [29], QR codes, RFID [25], and bar codes [30] or a fusion of these technologies [31]. Many of the researchers apply ML to different biometric features but an overwhelming majority have proposed attendance systems that are based on facial recognition systems. Although fingerprint and retina scan methods apply ML in their operations, their obtrusive nature makes them less attractive. Meanwhile, face recognition has the potential to enhance managing the attendance system, decrease mistakes associated with the manual attendance process by offering automatic and reliable tools, boost privacy and security aspects, prevent false attendance inputs, and deliver attendance reports [32], [33]. These systems use cameras to capture students’ faces and apply different ML algorithms to recognize students as soon as they enter the classroom. The attendance status is automatically updated in the database and instantly available to stakeholders. The face recognition problem can be subdivided into the face detection, the facial feature extraction, and the classification problems. The algorithms mostly used for face detection and feature extraction are Viola-Jones [34]–[36], histogram-oriented gradients (HOG) [37]–[39], PCA and LCA [40], [41], and local binary pattern histogram (LBPH) [42]. When it comes to classification, SVM is the most widely used ML algorithm for face recognition deployed for attendance systems [34], [35], [37], [38].

Another widely used algorithm for face recognition is LBPH, which combines the local binary pattern (LBP) and HOG techniques to improve the accuracy [7], [8]. It is known for performance and precision as well as its ability to identify an individual’s face from both the front and the side. Several researchers have applied this algorithm to achieve higher accuracy [43]–[46]. In previous study [43], the researcher utilized LBPH and Haar cascade algorithms for the tasks of face recognition and detection. The results demonstrated an impressive accuracy of approximately 96.68%, irrespective of lighting conditions. Notably, the study concluded that the performance of LBPH surpassed that of Eigenface and Fisherface algorithms, underscoring its superiority in these applications. LBPH achieved 100% accuracy in face recognition compared to only 73.3% by Eigenfaces [44]. The Haar cascade frontal face algorithm could successfully recognize the students based on an image captured within a classroom setting [45].

The review of the literature showed that recently many considerable ML applications in class attendance tracking used facial recognition approaches, which utilize the complex NN-based algorithms. convolutional neural networks (CNNs) are an example of a deep learning algorithm that is widely used in face recognition attendance systems due to their high accuracy, robustness to variations, and capacity to learn and extract meaningful features from images [37], [38], [47]–[50]. CNNs receive input from images and determine which features or objects ML can use to recognize and separate one image from another [51]. The system proposed in [47] uses NN (referred to as single-shot multi-box detectors) for face detection and visual geometry groups (VGG) networks for multi-class face recognition purposes. Single shot multibox detector NN model extracts the confident region exhibiting the human face features. Single shot detectors (SSD) use the WIDER dataset. This dataset contains coarsely annotated faces for different people in different situations. The WIDER dataset is comprised of 9,798 training images, 1,960 validation images, and 1,305 test images. The author claimed an overall accuracy of 94.66%. Research by Warman and Kusuma [49] applied deep learning CNNs and transfer learning for face recognition. The authors used three pre-trained CNNs, namely AlexNet, GoogleNet, and SqueezeNet, and trained them with a custom dataset of two hundred images classified into ten classes. All three networks achieved a validation of over 93%. CNN has been reported to be more accurate than the LBPH [52]. Most of the researchers used desktop computers as hardware platforms but some also used Raspberry Pi [53]–[55] as well as mobile [56], [57]. All the mentioned systems generate reports in the form of .xls, CSV, or XML format or directly update the information in the database. It should be noted that a single camera might not be enough to take the attendance of the class with a large number of students [48]. The future trend is integrating the ML-based attendance system with the LMS.
ML can offer important benefits for class attendance, yet there are also some challenges and potential drawbacks to consider. One of the important challenges is offered by the variability of factors which makes it hard for the algorithms to make accurate predictions under different scenarios and conditions. The quality of datasets also affects the performance of the system. A noisy dataset may degrade the performance of the system. An additional challenge is to find suitable algorithms that can be run on different devices with limited resources and be accurate at the same time. Acquiring and analyzing the attendance data using ML can raise privacy concerns, as it involves collecting and processing personal information. The accuracy of attendance data collected through ML can be affected by a variety of factors such as technology malfunctions or student behavior. Therefore, it is important to validate the accuracy of data collected and implement methods to address any errors or inconsistencies. Implementing an ML system for attendance tracking can require significant investment in technology, training, and maintenance. Smaller institutions may not have the resources to support such initiatives. ML algorithms can have biases and perpetuate discrimination if they are not meticulously designed and trained. Careful attention should be paid to the ethical considerations of using ML in class attendance and to ensure a fair and equitable system for all students. In some cultures, attendance may not be as important, or students may have other responsibilities that conflict with regular attendance. ML may not be able to capture or accommodate these cultural and practical nuances.

3.1.3. College/university selection
ML techniques can help in the challenging decision of university or program-of-study selection in several ways [58]. One way is to help high school students select the most appropriate field [59] and the right HEI to attend [60], as the majority of them are usually, unsure of which university and which major to pursue after high school [61]. Another way is to help the educational institution process the collected data regarding students’ performance [62] and use it to identify potential patterns and predict the student’s future performance while attending different programs of study. Moreover, ML can help educational institutions to lower the students’ dropout cases [63]. This is doable by developing accurate prediction models for identifying the student groups with dropout motives, detecting potential reasons for dropping out, sending early warning signals, and providing recommendations to develop individualized or group support approaches for the at-risk students.

Suggesting potential programs of study for high school graduates were investigated and used the RF algorithm to predict undergraduate majors for high school graduates [59]. The input to the model was related to the academic history of the student and the expected job market. The results showed that the proposed recommender system performed better than other published approaches and was able to achieve an accuracy level of 97.70%. The performances of RF and gradient boost classifiers (GBC) for predicting the students’ potential fields of study before admission [58], were as good as DT and extra tree classifiers (ETC) [64]. For deeper analysis, Mostafa and Beshir [60] studied which factors affect the students’ decisions to choose a university. The SVM was used in [65] and naïve bayes (NB) in [66] as classification algorithms. The findings showed that pupils with British-IG high school diplomas used an individual ecological system [67], the decisions of those with national high school certificates were affected by their exosystem [68], whereas the American high school graduates’ decisions were affected by their parents and relatives, including their microsystems. The NB algorithms were found to be better than SVM in accuracy, recall, and precision.

To find the most promising ML technique, Buraimoh et al. [62] compared six ML models for predicting student success in a technology-mediated learning environment. They have proved that students' behavioral attributes within a learning management environment are significant factors in forecasting their performance. The study showed, also, that the two ML models; classification and regression tree, and linear regression had achieved 86% accuracy, the best among the other models. Dropping out of the university has economic and social impacts, besides students’ negative self-perception due to failing to graduate. Segura et al. [63] attempted to mitigate such impact and problems using ML with data related to dropout after first year of study. They developed a feature selection process that was useable in identifying the variables correlated with dropping out. ML models (SVM, DT, and NN) and LR, were then utilized to predict the likelihood of dropping during or after the first semester. It was shown that dropout detection works with enrollment variables, and it can further improve after the results of the first semester of study.

3.1.4. Timetabling and scheduling
HEIs consistently rely on planning and controlling the scheduling of teaching sessions, labs, and classrooms to ensure efficient resource utilization and to meet the needs of both students and faculty [69]. Due to numerous constraints, parameters, and frequently abrupt changes in requirements, task scheduling is a well-known complex nonlinear process [70]. These restrictions or constraints are classified in two types; the first is represented by the ones with time restrictions, the ones with place restrictions (e.g., room allocation), and is known as strict/hard restrictions that cannot be avoided. The other type can be represented by students and faculty preferences and is known as soft restrictions which are manageable and can be optimized [71].

These constraints are important to be identified and classified carefully [72], for example as listed in Table 1, cause they are further used to formulate the scheduling problem, build the process model, select the appropriate algorithm, and they clearly affect the resultant solution [73]–[79].

To shed light on the research endeavors to offer successful solutions for this highly constrained problem non-linear, Table 2 provides an overview of the possible problem-solving options. This table reflects current methodologies, key characteristics, strengths, and shortcomings. The presented methods (algorithms) span a wide range of techniques from classical (e.g. graph coloring [73] to well-known optimization and search techniques [80] to advanced AI [81] and ML techniques (e.g. deep-belief networks [82]).

| Table 1. Soft and hard constraints for the university timetabling and scheduling problem |
|-----------------------------------------------|-----------------------------------------------|
| **Soft constraints**                         | **Hard constraints**                          |
| - Honors and general courses are preferable to be scheduled in non-overlapping time slots [73]. | - Courses with common students are not to be scheduled concurrently. |
| - The number of students enrolled in a course must fit in the designated class room [74]. | - The daily timetable has a maximum total of 8 hours of available periods. |
| - For a given curriculum, lectures should be sequential. | - In a given period, no more than one lecture can be assigned to the same room [74]. |
| - A minimum number of students must be registered in order for the course to start [75]. | - The same course must be assigned to different time periods. |
| - The lecturers' workload should be taken into account [76]. | - Courses by the same lecturers must be assigned to different time periods. |
| - Breaks must be scheduled between classes and prior to other classes. | - The schedule needs to be prepared in accordance with the university's academic calendar. |
| - There cannot be two reservations for lecture locations at the same time [78]. | - The professors' free time (such as office hours and resets) must be considered and considered. |
| - Every course needs to be paired with a location and time period [77]. | - The number of courses must not be more than the available rooms that can accommodate. |
| - Departmental courses must be placed in slots after faculty general courses [78]. | - It is necessary to plan the lecture places and spaces once. |
| - No student organization is hosting two events simultaneously. | - No student organization is hosting two events simultaneously. |
| - No lecturer is scheduled for two events at once. | - No student organization is hosting two events simultaneously. |
| - No event is held in a space that only accommodates fewer attendees than the actual attendance [79]. | - No event is held in a space that only accommodates fewer attendees than the actual attendance [79]. |

| Table 2. Comparison of different course scheduling problem algorithms |
|-------------------------------|-------------------------------|
| Approach                      | Major features               | Strengths                  | Weaknesses                  |
| Graph coloring [73], [83]. Tabu search [84], [85]. | Obtaining the best solution. Preventing any repetition and repetititon of operations. | Avoid local optimal. Not to be in a state of local optimality. | The solution has a major impact. Extensive execution time. |
| Simulated annealing [69], [86]. | Make use of the temperature factor. Acceptance of a worse program is a possibility. | Capability to break free from the local maximum. Provides an optimal solution to the best of its ability. | There is no such thing as a greedy performance. Impractical because of the large number of possible solutions. |
| Random repeat optimization algorithm with hybrid neighbors [83]. | Requires a sizable chunk of memory. | Flexibility in dealing with local optimization problems; high convergence rate. | Early convergence and cohesiveness in local optimization, as well as population diversity reduction. |
| Particle swarm optimization algorithm [87]. | | | |
| Hill climbing algorithm: the first choice [88]. | Appropriate for problems requiring a large number of great functions. | High performance, low execution time, and low memory consumption. | Not perfect and not optimal. |
| Genetic algorithm [69], [70], [89]–[91]. Bee's and Bee colony algorithms [69], [88], [92]. | Initiated from any random state. Local and general research are available and coexist. | Fast convergence and there is no need for advanced mathematics. High efficiency in reaching the best or optimal solution. | No utilization of information distribution. Use the coordinate number of variables; dependent parameters can be defined. High storage capacity and simulation time. |
| Memetic algorithm [93]. | Local search. | Capable of solving problems with topical solutions. | Tendency to produce inaccurate outcomes as the environment grows and the number of layers increases. |
| NN [91], [94], [95], [65]. | Inspired by the biological nervous system for data and information processing. | Effective in dealing with complex and large problems. | |
| Sequence-based Selection Hyper-Heuristic (SSHH) approach [36]. Deep-Belief Network (DBN) [82]. Reinforcement learning [96]. | Sequences of innovative methods are chosen and used. Learn the constraints and inputs in one step. Hyper-heuristic approach with the cooperative scheme. | Discovering the inherent characteristics and relationships in data and generalizing them to other situations. Better score, less error, and execution time compared to other methods like the SSSH approach. Flexibility to adapt to exam timetables with distinctive characteristics. Little need for any parameter tuning. | Long execution and runtime. Needs an objective function that is difficult to formulate. The training and learning procedures are cumbersome. |

Applications of machine learning in operational aspects of academia: a review (Muhammad Nadeem)
Even though many approaches have been investigated for the timetabling and class scheduling problems, a comparative study among all these techniques is still not well covered in the literature, which is highly recommended for future study. This comparative study should analyze and evaluate these approaches thoroughly and justify the superiority of one over the other heuristic-based or mathematical-based ones. Moreover, more investigation is required to add additional features to the surveyed approaches [97], as well as more analysis is highly recommended to improve their performances. Also, more tests and applications of these approaches need to be implemented to further prove their efficacy. These approaches proved successful in this application (classroom scheduling), however, using the same techniques has the potential to benefit other industries as well that depend on scheduling in their operation, such as healthcare and transportation.

3.1.5. Student loans and financial aid

Many countries around the world offer student loans for those pursuing higher education to help them bridge the financial gaps, and to address educational needs. Governments offer various kinds of incentives such as low-interest and prolonged repayment plans to promote this practice. But the number of students defaulting is increasing resulting in billions of dollars being put at risk. Therefore, it is important to identify students who can default and strengthen monitoring. The allocation of scholarships is crucial to both institutions and students who want to enter universities. The recent hike in prices resulting from the prolonged closing of business activities due to COVID-19 has also resulted in inflation in college fees, making it too expensive for many students, and threatening a decline in student enrollment.

Scholarship allocation is a complex task that involves evaluating multiple factors such as academic performance, extracurricular activities, financial need, and community involvement. University scholarship plans are not always governed by rules. The business procedure for choosing the candidates for rewards may not be effective and efficient in these circumstances. To minimize business process inefficiencies, existing work has shown that data mining methods can be useful for developing rule-based scholarship allocation policies. ML can be used to automate the process of allocating scholarships and ensure that scholarships are allocated fairly and efficiently.

The first step in using ML for scholarship allocation is to collect and prepare the data. This involves gathering information about the applicants, including their academic records, financial status, and personal background. The data should be cleaned and organized to ensure that it is accurate and complete. Ma [98] applied the LR to identify the nine significant influencing factors, and eight common factors were extracted through factor analysis which can explain 80% of all of the original variables. Rohman et al. [99] demonstrated how an ID3 DT algorithm could produce guidelines to choose the scholarship candidates who stand the best chance of receiving one. This basic principle made it possible to identify applicants quickly imparting the scholarships. Alhassan and Lawal [100] applied tree-based models for the binary classification of students in Nigeria for their eligibility for scholarships for doctoral studies. The system was trained using 800 training cases each having ten attributes including grades, qualifications, and status. Xin Hao [101] automatically evaluated student scholarships and exclude the factors influenced by humans. The dataset included student data and assessment results of 671 students of them 71 test students were used to train SVM models.

Several studies have combined enrollment prediction models with optimization techniques for the purpose of recommending financial aid strategies. In 2015, Sarafras et al. [102] used a genetic algorithm to discover a scholarship strategy that maximized overall enrollment after using a NN model to predict enrollment. Sugrue [103] created a merit-based aid allocation model using data from the University of Miami and employed LR to predict enrollment and a linear programming model to optimize the standard of the incoming class. Aulck et al. [17] applied ML and mathematical optimizations on data mining in an application by a large American public university for optimized disbursement of merit-based scholarships to students to maximize enrollment. The authors first developed the model of predicted enrollment and then applied the genetic algorithm for optimizing the scholarship disbursement. They suggested that smaller scholarships result in increased enrollment. Phan et al. [104] proposed an interactive feedback loop comprising a gradient-boosting classification algorithm and auto-start local search optimizer which predicts enrollment and revenue for a given strategy for award and aid allocation by using the classifier. Optimizer uses these predicted outcomes to improve the current strategy meeting institution-centric and student-centric objectives. There exists an interactive feedback loop between the classifier and optimizer. Gradient boosting and extreme gradient boost exhibited the highest performance. The system is aimed at enhancing enrollment and revenues which are pure institution-centric objectives. This model was further improved by adding student-centered objectives such as affordability [105] and using simulated quenching to optimize the strategy for optimizing merit-based or need-based scholarships. They were able to identify different strategies to improve accessibility to university education without compromising other objectives.

However, it is important to note that ML models are not a panacea for scholarship allocation. They can help automate the process and reduce bias, but they are only as good as the data and algorithms used to
train them. It is also crucial to ensure that the model is transparent and interpretable so that decisions can be explained and challenged if necessary. Human oversight is still necessary to ensure that the model is fair and just and that deserving applicants are not overlooked. Finally, the student and financial information required for these forecasts and optimizations is kept in the university database and it is not easy to have access to this data due to multiple reasons such as poor infrastructure or lack of staff expertise. Universities have access to this data but are slow to use it for their own purposes, possibly because of inexperienced staff or inadequate data infrastructure [16].

3.1.6. Student project allocation

Senior students are required to carry out a graduation project as an important component of the undergraduate degree program. The graduation project provides an opportunity for senior students to “prove” and demonstrate their proficiency in their respective fields by applying the knowledge and skills they have acquired throughout their academic journey. Therefore, the allocation of a graduation project is a very important and critical process that requires careful planning.

The student project allocation (SPA) problem is a well-recognized nonlinear discrete optimization problem whose goal is to compute an optimal or near-optimal allocation that meets various constraints and preferences of the students and the supervisors. The origin of the SPA problem dates back to the seminal paper of Gale and Shapley [106] where the stable marriage (or machine) problem (SPM) was introduced. In a scenario with N men and N women, each individual possessing a preference list of potential partners from the opposite gender, the challenge is to create marriages (matches) between the men and women in a manner that ensures no pair of a man and a woman would both prefer being matched with each other over their current matches. When there are no such pairs of people, then the set of marriages is defined as stable. Their “stable matching” algorithm is based on the idea of “deferred acceptance” and always yields a stable matching. This algorithm has had a significant impact on many real-world applications in economics, education, and computer science. Where two groups of individuals with different preferences need to be matched together in such a way that all pairs are stable. Lloyd Shapley and Alvin Roth were awarded the Nobel Memorial Prize in Economic Sciences in 2012 “for the theory of stable allocations and the practice of market design”. Making an analogy to the SMP in [106], the students and projects correspond to the men and women, respectively, in the SPA problem. Adding other factors such as the academic levels, skills, and interests of the students, the requirements of the supervisors, and available resources into account, then the SPA problem turns into an even more challenging problem.

The SPA problem can be put into two groups: those graduation projects done in groups (of 2 to 5 students), and those done individually. Some universities prefer graduation projects to be done in groups due to various reasons [107]–[109]. This has the potential to substantially lessen the burden on supervisors and simplify the SPA problem. Anwar and Bahaj [110] proposed an alternative approach where up to three students can form their own group for the graduation project. Hussain et al. [111] put the project allocation methods into different categories: those which are based on the preferences of both students and lecturers, project selection by students based on project titles, those based on supervisors and/or project category, and based on students’ own proposals. They present a detailed review of the project allocation methods in the literature. For a survey on the SPA, see e.g. [112]. Thanks to the fruitful results developed by the SPA research community since the 1960s [113], the SPA process has been automated by various universities around the world [108]–[110], [114]–[118].

However, the use of ML in the automation efforts above is not observed. The closest research work to the SPA problem was an automated thesis supervisor allocation process using ML has very recently been presented in the research [118], where the authors use a DT Classifier in Python for the training of their classification model, as human experience can be translated to a DT. The use of AI/deep learning techniques for solving complex SPA problems is an attractive research area, which is worth investigating. The problem as defined lends itself well to the ML approaches. However, because the subject has not been explored yet in literature, in what follows we summarize how the SPA problem has been solved using different methods and approaches before.

The student project allocation problem is typically solved using mathematical optimization or heuristic algorithms. Various methods and algorithms have been proposed previously, including methods based on preference lists [106], matching algorithms, game theory [119], [120], genetic algorithms [121], integer linear programming [122], and fuzzy logic systems [123], [124]. The study by Yahaya et al. [109] used constraint optimization techniques and developed a Java platform that mainly focused on students’ preferences for projects only, to guarantee that a maximum of students get their first-choice projects. A dataset containing 95 students and 12 supervisors with 108 project titles in the Computer Science program of the Department of Mathematics, Usman Danfodiyo University is used in their study [109].

There are many variants of the SPA problem in which the goals are to achieve stability, fairness, and transparency, taking into account the complex and diverse preferences and constraints of the students and...
supervisors [125]. For example, in 2007, Abraham et al. [126] defined the SPA-P problem, an extension of the well-known hospitals residents with ties (HRT) problem. The hospitals/residents problem has been a focus of research where the goal is to compute a stable matching of doctors (or residents) to hospitals founded on their preference lists [120]. Teo and Ho [107] described a structured approach to the implementation of the SPA in an electrical engineering undergraduate program including tips on how to organize teams and how to allocate resources. They presented a computer program in C to allocate projects yielding a possible viable solution but not necessarily an optimal one.

Anwar and Bahaj [110] presented two integer program models to solve the SPA problem with several constraints. The SPA with preference over students was studied by several researchers [126], [127]. According to their methodology, preference lists are submitted by students over projects that instructors have presented, and preference lists are provided by faculty over students who express interest in one or more of their projects. El-Atta and Moussa [128] present a method to create a general SPA where faculty have priority lists over pairs, and the candidates have priority lists over projects.

Some of these methods were later applied to the data from the Department of Civil and Environmental Engineering, University of Southampton, for the academic year 2001–2002. Research by Pudaruth et al. [108] presented a framework for the SPA in the Computer Science and Engineering Department at the University of Mauritius. The system allocates projects to students to maximize the number of students who get their first choices in their preference list and to balance the load of faculty as well. Kwanashie et al. [116] proposed a model for the SPA problem based on individual or group projects offered by a supervisor. They employed a greedy method to model the SPA as a network flow problem, aiming to maximize the number of students assigned to higher-ranked projects. This approach led to achieving the maximum optimal matching or assignments.

There are quite a lot of variants of the original SPA problem. Manlove et al. [129] proposed the so-called SPA with preferences over projects (SPA-P) where faculty orders a subgroup of projects rather than students. They proposed an approximation algorithm for searching for a stable matching where the number of students assigned to their favorite projects is also maximized taking the constraints into account. They proposed an integer programming approach to solve the SPA-P [129]. Some other variants are the SPA problem with faculty preferences over students (SPA-F), the SPA problem with lecturer preferences over students with ties (SPA-ST). Three distinct ideas of stability—weak stability, strong stability, and super-stability—are produced by the preference lists having ties [125], [130]. A thorough mathematical investigation of these SPA variants is presented in previous study [125].

The problem of finding a maximum size stable matching, designated as the MAX-SPA-P problem, which is another variant of the SPA problem, has recently been a focus of research as the theory of stable matching in the MAX-SPA-P problem finds extensive applications in education and economics. This includes establishing matches between buyers and sellers, suppliers and carriers, students and universities, among other scenarios involving preference list constraints and the need for optimizing the maximum number of matchings. Nguyen et al. [131] present a heuristic search method for solving the MAX-SPA-P problem of large sizes. The MAX-SPA-P problem is also NP-hard [125]. Olaosebikan [125] presented an integer programming model to enable MAX-SPA-P to be resolved optimally in the general case in his Ph.D. thesis. Recently, Viet et al. [132] proposed a local search strategy for solving the SPA-P of large sizes. They proposed a heuristic algorithm, called SPA-P-MCH, based on the min-conflicts algorithm to solve the MAX-SPA-P problem of large sizes.

Taking all the complex and diverse preferences and constraints of both the students and supervisors and the resources, the SPA can therefore be seen as a complex multi-objective optimization problem. We think that this complex SPA optimization problem could and should be solved by contemporary AI/deep learning techniques and tools. Considering the latest developments in deep learning and AI, we envisage that the contemporary and unprecedented AI techniques could do much more than those of the standard methods (constraint optimization, integer programming, and dynamic programming) for optimizing the complex SPA problem soon taking also the following aspects into account:

- A large amount of student-specific data (e.g., all exams, course project reports, answers in exams, and social or cultural club activities) is generated for each individual student until they are senior students. This student-specific big data can be used to analyze the history of the student’s performance, project preferences, and supervisor preferences to identify some characteristic patterns and then to predict future outcomes. This can provide some meaningful insights when optimizing the SPA problem.
- Many students have difficulty choosing the most suitable senior project for themselves. Natural language processing (NLP) tools and Chabots can help them better understand and evaluate their preferences and choices to ensure that the SPA is based on accurate and reliable data.
Artificial intelligence can be used to generate multiple potential allocations and evaluate them based on various criteria, providing a few different alternatives for the student. In other words, the AI will offer an intelligent decision support system for the students in making up their minds about the project proposals. We should make sure that AI tools are used ethically and responsibly throughout the SPA process.

3.1.7. Curriculum development

The curriculum development problem refers to the challenges of designing, implementing, and evaluating a “high-quality” curriculum and evaluating its effectiveness. Some major challenges include aligning the curriculum with the student learning outcomes, improving the limited resources (laboratories, academic personnel, and technological infrastructure) in developing a comprehensive and effective curriculum, identifying and determining the outdated or irrelevant parts of the course content, linking the relevant courses, developing assessment methods to evaluate an academic program and to improve the effectiveness of the curriculum overall.

The curriculum development problem has been an important focus of research after Macdonald and Walker’s book in 1976 [133], where several recommendations were presented to reform the curriculum. Examining the history of the British education system, they argue that the traditional curriculum should be reformed because it became “old-fashioned” by focusing on scientific subjects and lacked relevance to the needs and expectations of contemporary students and society [133]. Later, Kessels [134] presented similar arguments and proposed a relational curriculum design with three key principles: integration, coherence, and relevance. The relevance part of the curriculum design implies developing the curriculum that relates to students’ lives and real-world experiences in their societies and the ecosystems in a broader sense.

Defining a curriculum is not a trivial task and thus there is not a consensus on its standard definition. Accordingly, various definitions have been proposed in the literature. Glatthorn [135] defined six curriculum types (recommended, written, taught, supported, tested, and learned). Later they provide step-by-step practical guidance for educators seeking to renew their curriculum in a systematic and collaborative approach [136]. Clements presented a framework about how to build “research-based curricula” [137]. A work-related curriculum is examined [138]. To address all these curriculum development challenges, developing a quality curriculum requires a collaborative, systematic, and iterative approach taking various inputs from a range of stakeholders like students, faculties, educators, and industries. into account. Almost the whole of the curriculum design community agrees that the traditional curriculum should be “reformed/renewed” taking the needs and expectations of all the stakeholders of our contemporary “digital era” into account.

During the last decade, unprecedented developments in the ML and AI areas opened new doors to curriculum developers in solving various problems. Rawatkal [139] proposed data mining the student records to determine curriculum structure. The same author deduced a curriculum in [140] using the methods outlined in [139]. The author proposes tree-based ML algorithms on historical records of students to prioritize the prerequisites of courses. He also proposes a machine-learning-based framework for the analysis of curriculum design. The availability of the “big data” of students especially after the 2016s made it possible to apply the modern ML techniques to the curriculum design problem. Ball [10] proposed refining the curriculum using ML to meet student needs. They first examine the attributes and academic needs of the student by using ML algorithms, and then accordingly apply essential modifications to the curriculum to boost graduation rates. The authors compare four different ML algorithms: DT, LR, AdaBoost DT ensemble, and RF to identify areas where students struggle in computer science majors. Finally, they propose changes to both curriculum and faculty mindsets accordingly.

Somasundaram et al. [141] propose to design a curriculum to specially meet the industry and job market requirements of a given area using the artificial neural network (ANN) methods. They use the backward propagation algorithm to train the ANN defining the job roles as the final goal and taking the necessary knowledge, needed skills, and attributes as prerequisites [141]. The courses and their learning outcomes were mapped to the program outcomes and then to the job roles. In other words, the series of courses and their courses as well as the required skills and competencies are modelled as an ANN, which is then trained by the back propagation algorithm [141]. In their simulations, they examine the Anna University curriculum and the national programme on technology enhanced learning (NPTEL) online courses, which were initiated by seven Indian Institutes of Technology. Similarly, Ketamo et al. [138] argue that keeping a curriculum current and using language appropriate to the workplace is one of the main design problems of a work-related curriculum. They also report that every curriculum is crafted using academic language, potentially lacking terms that hold the highest value for students actively searching for employment opportunities. Somasundaram et al. [141] build collaboration with the Helsinki Metropolitan Area Universities of Applied Sciences (3AMK) and Headai Ltd. and apply cognitive AI, big data, and NLP to build a real-time understanding of skills, competencies, knowledge, and abilities that workplaces seek. Georgiopoulos et al. [142] propose an integrated research and teaching model for incorporating the current ML research topics into the undergraduate curriculum at the University of Central Florida. The modules...
showcase contemporary machine-learning topics by using them as examples of applications that students previously learned in their courses [142].

In summary, any curriculum has some room for improvement. ML and AI provide tools to help reform or renew the curriculum taking the contemporary needs and expectations of the stakeholders like universities, students, faculties, educators, and industries, into account. ML and AI can be used all the way from the start point like analyzing the text of a curriculum by NLP techniques to improve the clarity and accessibility of the curriculum to the endpoint like determining a series of specific courses from the curriculum to provide a personalized learning path for an individual student based on his needs, interests, and learning styles. The literature review indicates that ML has far-reaching implications in the non-teaching aspect of academia; the tasks which do not focus on learning outcomes. Time-consuming administrative operations like data entry, scheduling, and record-keeping can be automated with ML. This boosts operational effectiveness by enabling personnel to concentrate on higher-value tasks like strategy planning and student support services.

3.2. Common challenges
3.2.1. Data quality

Although ML offers many benefits, it has its own challenges. ML needs a large amount of excellent-quality data for learning and making accurate predictions. There is strong evidence that predictions using ML can be prone to error if data is not handled properly [143]. The first is deciding the right amount of data for training the model since accuracy is highly dependent on that. Failing to do so creates biases and the system will not make predictions accurately. On the other hand, using a large amount of data to train the system will require a significant amount of time which may hinder some tasks of an urgent nature. Therefore, tasks should be planned carefully to avoid interruption in the function of the institution.

3.2.2. Lack of expertise

Although this data is available to the universities in the raw form, they lack the resources and expertise to make effective use of this data. One option is to give access to a third party that has expertise in handling the data, but this raises privacy concerns and universities are reluctant to give access to the sensitive third party. Another option is to train the staff, which is time-consuming and poses a continuous threat as the trained staff may leave. However, it is expected that universities will be left with no option but to open more to reap the benefits offered by this emerging technology. Also, the availability of a workforce in the future may encourage universities to train their staff without any fear of losing them. Furthermore, data should be collected carefully so that it is free of biases. If biases are not mitigated rigorously, they may be propagated and even magnified resulting in unfair and inaccurate outcomes [144].

3.2.3. Algorithm selection

A plethora of ML algorithms exist, each having its own pros and cons. There is no one solution fit to solve all problems. To make accurate predictions it is important that algorithms are selected carefully.

3.2.4. Transparency and accountability

One of the major areas of concern arising because of the application of the ML is related to transparency and accountability because systems can make wrong decisions and we cannot blame machines. Often the user is not informed that the decision process is automated via ML algorithms and presents challenges in explaining decisions, which affects trust and accountability. The end user is often left unsatisfied and frustrated. This becomes more significant when dealing with matters which affect students directly such as allocation of financial aid and project allocation. Therefore, a system should make clear when decisions are made automatically and when they are made by humans to promote openness. If possible, a decision or decision procedure should be justified. Similarly, if something goes wrong who will be held responsible and pay for damages? It is critical to ensure that a system is tested thoroughly, and risk analysis is performed to avoid anticipated harm before it is released. To date, there are no clear answers to these challenges. More legislation and policy recommendations should be drafted, and the reinterpretation of already-existing legal and regulatory frameworks is necessary.

3.3. Application specific challenges

Machine learning is a powerful technology with a wide range of applications across various domains. However, it also comes with a set of application-specific challenges that must be addressed to ensure its effective implementation in specific contexts. Here are some challenges specific to the application considered in this research.
Applications of machine learning in operational aspects of academia: a review (Muhammad Nadeem)

3.3.1. Predicting enrollments

One of the issues faced when predicting enrollments is the transferability of the algorithms as if developed for one scenario, it becomes hard to apply in different situations that require customization. For example, the model developed by Wu et al. [145] for predictive student enrollments takes into account the number of different departments a student can apply. If this constraint is relaxed and students are allowed to apply for more departments, then this might not work well. Seasonal patterns and changing trends can impact accurate forecasting therefore model maintenance is necessary which should be updated to incorporate these factors. Similar challenges are also faced by students when selecting universities or colleges. Selection criteria for universities may alter over time to reflect shifts in the job market, societal values, and educational objectives. It can be difficult to modify ML algorithms to account for these shifting standards.

3.3.2. Student class attendance

Machine learning for facial recognition has come a long way, but there are still many obstacles to overcome. Facial recognition is susceptible to spoofing as attackers can utilize images, movies, or 3D masks to mislead the system. It can be difficult to provide informed permission and make sure users are aware of use cases for facial recognition. Users should have the choice to opt-out and be informed of how their facial data will be utilized. This makes it hard to deploy such systems. System reliability decreases in real-world environments with changes in lighting, position, facial emotions, and occlusions [52].

3.3.3. College and university selection

Selecting the right college or university involves numerous diverse and subjective factors, presenting challenges for software and ML models. These challenges include the need to account for a wide array of criteria such as academic reputation, location, and financial considerations, while also adapting to dynamic changes in student preferences and the evolving educational landscape. Ensuring the quality and availability of data, addressing biases, building user trust, and considering ethical implications are crucial aspects. Additionally, the unpredictable nature of future trends, limited historical data for some institutions, and the complexity of incorporating long-term success metrics pose further obstacles. The multidisciplinary nature of this problem requires collaboration between experts in ML, data science, education, and ethics to develop recommendation systems that are transparent, fair, and trustworthy.

3.3.4. Timetabling and scheduling

Timetabling and scheduling using ML are useful techniques that are normally treated as an optimization problem, but they come with many challenges. Timetabling and scheduling problems can include a lot of variables and constraints, which can create high-dimensional optimization spaces and make it difficult and computationally expensive to identify the best solutions. It is not easy to define an adequate objective function that strikes a compromise between competing objectives, such as avoiding classroom conflicts while boosting student preferences. The exam schedule is one of the processes that interfere with making decisions, and frequent dependence on course scheduling makes the optimization challenge even harder [146].

3.3.5. Financial aid and student load

When using machine learning for financial aid and student load, one of the major challenges faced is the amount of limited data available. Unlike other applications of ML such as facial recognition where existing datasets may be augmented, one needs to wait for the data to be produced in real-time. When data frequency is high, it is considerably noisy leading to unreliable models. Data evolution is another difference between finance and other fields where ML is used. Using picture identification as an example once more, human photos always contain similar qualities. ML techniques can recognize images using these attributes. In contrast, financial data, like the financial markets, develops and changes over time. As a result, it is challenging to think that financial variables now have the same significance as they had decades ago [147].

3.3.6. Curriculum development

Developing a curriculum is a multifaceted challenge for software and ML models considering the diverse learning styles of students, the rapidly changing landscape of industry needs, and the necessity for personalized, interdisciplinary learning paths. Integrating emerging technologies, designing effective assessment methods, and considering cultural inclusivity add further complexities. Resource constraints, faculty training, and ethical considerations must be carefully navigated, and the challenge of keeping the curriculum up to date with technological advancements is ever-present. Ultimately, successful curriculum development through software or ML requires a holistic approach that combines educational expertise, technological innovation, and collaborative efforts between educators, industry professionals, and technology experts.
3.3.7. Project allocation
When allocating projects to students’ various restrictions, including project availability, supervisor preferences, and project prerequisites, may be present. It can be challenging to include these constraints in a machine-learning model. Over time, students’ preferences for project topics can shift, and fresh assignments might become accessible. Models for ML must change to reflect shifting preferences and project options. It can be difficult to define the optimization goals for project allocation (such as enhancing student satisfaction or reducing supervisor burden) and convert them into ML goals [111].

3.4. Future directions and potential
The exploration of ML applications in the administrative and operational aspects of academia has already yielded significant insights and opportunities. However, the field is poised for continued growth and evolution, offering a range of exciting future directions and untapped potential. Here are some key areas that warrant attention in the coming years.

3.4.1. Enrollment prediction
Machine learning enrollment prediction is a dynamic and ever-changing field. In the future, enrollment forecasts will be more precise and complicated thanks to advancements in ML algorithms, including deep learning methods like NN and transformers. The detailed patterns in enrollment trends can be captured by these algorithms, which can handle high-dimensional data. These models will be integrated with predictive analytic tools for making admissions in other areas. Furthermore, predictive models will not only be used for global enrollment trends but can also be customized to meet the unique requirements of institutions. Universities may collaborate in the future by sharing enrollment information and insights while protecting student privacy and data security facilitated by blockchain-based technologies and federated learning [148].

3.4.2. Class attendance
Future facial recognition systems will improve their accuracy under tricky situations like dim illumination, occlusions, (such as masks), and fluctuations in position or facial expression [149]. The 3D facial recognition and liveness detection techniques will become more sophisticated and increasingly applied to counter spoofing [150]. Future breakthroughs in the application of ML to class attendance analysis are anticipated to be substantial. Institutions will create smartphone applications and chatbots that work with ML models to send students individualized reminders and rewards for showing up to class. In order to improve the precision of the attendance system and avoid false positives/negatives, future systems will combine facial recognition with other biometrics, Wi-Fi connections, and geo-location [52].

3.4.3. University selection
There are several fascinating trends and advances in the future of university selection utilizing ML that could completely alter the admissions procedure. Applicants will receive tailored university recommendations from ML algorithms based on their academic accomplishments, interests, extracurricular activities, and career goals. Students will find the top institutions with this information. ML will provide a holistic application evaluation beyond academic measurements by examining a wider range of attributes, such as soft skills, leadership traits, and creative accomplishments [151]. ML-driven chatbots will assist candidates through the application process by providing them the information, direction, and real-time support, enhancing the candidate’s experience [62]. These chatbots can even help in conducting and analyzing the interviews, helping universities assess the communication skills and personal traits of the applicant. Also, more sophisticated NLP algorithms will be employed to gain a deeper knowledge of applicants’ motives, objectives, and communication skills by analyzing and comprehending application essays and recommendation letters [152]. Using a hybrid model involving both human judgment and ML may ensure that decisions made are well-informed, fair, and aligned with institutional values.

3.4.4. Student loans and financial aid
The use of ML for student loans and financial aid is developing quickly, with several upcoming developments influencing the market. To maximize aid distribution, the institutions will be able to design highly individualized financial aid packages for students considering each student’s financial condition, academic performance, career objectives, and other pertinent aspects [153]. Alternative credit scoring models will be developed using ML that consider non-traditional data sources such as education and employment history beneficial for the student who has little credit history [154]. Furthermore, the model based on student performance and engagement can be integrated directly with the LMS to identify deserving students.
3.4.5. Timetabling

The future of developing software and ML models for timetabling involves various promising directions. This includes the integration of sophisticated optimization algorithms, ML for predictive analytics and pattern analysis, dynamic and adaptive scheduling systems, and multi-objective optimization to balance conflicting objectives. One of the issues faced is the lack of publicly available datasets and there is a need to ensure that such datasets are made public in an easy-to-use format so that researchers may reproduce the results and benefit from it [155]. Cloud-based solutions, internet of things (IoT) device integration, and user-centric design aim to enhance collaboration, resource sharing, and user experience. Blockchain technology may contribute to transparency, while real-time collaboration features and continuous learning models ensure adaptability and ongoing improvement. Considerations for energy efficiency and a holistic approach that involves collaboration between academia, industry, and technology experts are pivotal in shaping innovative and effective timetabling solutions for educational institutions [156].

3.4.6. Student project allocation

The future of software and machine learning models for SPA holds substantial potential. Skill-matching algorithms can optimize the process by assessing students’ skills and matching them with project requirements. Preference learning models enhance personalization, while a dynamic project database, collaboration with industry, and efficient resource utilization contribute to relevant and practical project allocations. Feedback mechanisms, adaptive learning models, and the integration of soft skills assessment ensure continual improvement and well-rounded project teams. Predictive analytics for project success, blockchain for ownership and contributions, and real-time communication platforms further enhance transparency, fairness, and collaboration [157]. By embracing these directions, these models can revolutionize the project allocation experience, fostering a more personalized, efficient, and rewarding engagement for students. Collaboration between educational institutions, industry partners, and technology experts is key to unlocking the full potential of these advancements.

3.4.7. Curriculum development

The future of developing software and ML models for curriculum development is poised for significant advancements. ML can contribute to personalized learning paths by analyzing individual student data, and dynamic and adaptive curricula that respond to real-time industry trends [158]. Predictive analytics can forecast skill demands, while interdisciplinary integration and automated content creation enhance the holistic educational experience. Continuous evaluation and feedback, collaboration platforms, and ethical curriculum design are crucial components. Integration of emerging technologies, blockchain for credential verification, and adaptive assessment strategies further enrich the potential of these models. The collaborative efforts of educators, industry professionals, and technology experts will be pivotal in harnessing these capabilities to create more responsive, inclusive, and innovative educational programs. As ML continues to advance, it holds immense promise for the transformation of administrative and operational aspects of academia. To realize this potential, researchers, educators, and policymakers must collaborate in addressing challenges, promoting ethical use, and pursuing innovative solutions. The future of academia is intertwined with the future of ML, and the possibilities are as vast as the potential of the technologies themselves.

3.5. Limitations of the study

One of the limitations of this study is its complete reliance on secondary sources such as published papers or reports. This is prone to mistakes or misinterpretations in these secondary sources, which may result in an unintentional propagation of inaccurate information or biases. Another limitation of this research is the absence of quantitative testing which backed by statistical analysis may enhance the generalizability of this study [159].

4. CONCLUSION

This paper furnishes a contemporary overview of the utilization of ML in the administrative domains of academia, exposing associated challenges. Through the comprehensive review undertaken, a majority of the university’s supporting and administrative functions have been subject to investigation employing ML methodologies. Spanning from class attendance management to augmenting administrative efficiency, ML applications present diverse opportunities for constructive transformation. Nonetheless, it is imperative to recognize and address the persistent challenges inherent in this transformative trajectory. Noteworthy among these challenges are data privacy considerations, ethical implications, and the imperative for robust infrastructure. The study reveals that ML has already found widespread application in various administrative facets of academia. Furthermore, there is a noticeable inclination toward expanding its implementation. The increasing availability of ML-ready tools and libraries has contributed to the growing
popularity of ML. Consequently, institutions are prompted to engage in campus-wide deliberations regarding the impact of ML on administrative processes. The dynamism of exploring ML in academia is evident, presenting numerous promising prospects. By apprehending the potential of ML and efficiently navigating challenges, the academic community can optimize administrative and operational processes while fostering innovation and advancements in education. As academia adapts and progresses, ML is poised to play an integral role in shaping its future.

REFERENCES


Applications of machine learning in operational aspects of academia: a review (Muhammad Nadeem)