ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #40839

Literature Review of Analyzing and Predicting Students' Performance in Examinations

Sakhi Aggrawal, Purdue University at West Lafayette (PPI)

Sakhi Aggrawal is a Graduate Research Fellow in Computer and Information Technology department at Purdue University. She completed her master's degree in Business Analytics from Imperial College London and bachelor's degree in Computer and Information Technology and Organizational Leadership from Purdue University. She worked in industry for several years with her latest jobs being as project manager at Google and Microsoft. Her research interests include: workforce development, engineering education, project management and teamwork. Her current research focuses on integrating project management processes in undergraduate education. Her main goal is to understand how work management and product development practices widely used in industry can be modified and adapted to streamline undergraduate STEM education.

Literature Review of Analyzing and Predicting Students' Performance in Examinations

Abstract

Background: Student dropout continues to be a critical problem in education. The sooner students at risk of dropping out are identified, the sooner necessary measures can be taken to support and guide them. Schools and universities are implementing data science methods to analyze available data, identify patterns, and extract information to support decision-making and effective student learning. However, student learning is often difficult to assess. Examinations are a popular assessment tool for testing students' knowledge, skills, and aptitude among others. Student performance in examinations can indicate their risk of dropping out, therefore, it becomes critical to analyze and predict their examination performance.

Purpose: In this literature review, we focus on exploring and analyzing existing work in this field to develop fundamental domain knowledge and avoid duplicative work. The specific research question was: What types of knowledge already exist in relation to analyzing and predicting students' performance in examinations?

Methods: The steps followed for performing this literature review were: (1) identifying the scope and research questions, (2) defining the inclusion and exclusion search criteria of literature, and (3) classifying and cataloging the literature sources that relate to analyzing and predicting students' performance in examinations. The final data set is comprised of a total of 10 papers that meet our criteria.

Results: Our literature review reveals that the field of analyzing and predicting students' performance in examinations is still in its nascent stages. While researchers have been developing and testing models for three decades, few studies have applied and verified their research on a larger scale. We observe a significant potential for developing intelligent systems, particularly in the context of learning management systems (LMS) and intelligent tutoring systems (ITS), which gained prominence during the COVID-19 pandemic. However, the exploration of qualitative aspects of student learning behavior remains limited, highlighting a need for further research in educational data mining (EDM). Additionally, while existing research has primarily focused on predicting performance in subjects like computer programming, STEM, and English, there is an opportunity to extend these prediction models to non-STEM subjects, such as business management, finance, sociology, psychology, hospitality management, nursing, and fashion, potentially revolutionizing admissions and hiring processes in these fields.

Implications: The synthesis of research findings highlights the importance of striking a balance between algorithmic predictions and humanistic considerations in education. The integration of data-driven insights into the learning experience, through methods like Intelligent Tutoring Systems and online platforms, presents promising avenues for personalized education. However, ethical concerns surrounding data privacy and algorithmic bias must be addressed to ensure equitable educational opportunities. Additionally, the identified research gaps, such as cross-course validation and intelligent tool development, offer opportunities for future exploration. As predictive models mature, their successful integration into educational practices depends on rigorous testing and refinement to provide actionable insights to educators, administrators, and students in real time.

Keywords: student dropout, data science, predictive modeling, examination performance, educational data mining, personalized education, algorithmic bias, intelligent tutoring systems, engineering education

1. Introduction

The urgency to address student dropout rates has never been more evident. In 2020, the U.S. Education Department's National Center for Education Statistics [1] reported that 2.0 million students between the ages of 16 and 24 dropped out of high school, translating to roughly 5500 students every single day. This brings the overall status dropout rate to a concerning 5.3 percent. Similarly, college dropout rates were a whopping 32.9% in 2022 [1]. These statistics highlight the importance of understanding students' reasons for leaving educational institutes as the sooner students at risk of dropping out are identified, the sooner necessary measures can be taken to support and guide them.

Schools and universities are increasingly turning to data science methods to comprehend this phenomenon. By analyzing educational data, these institutions aim to identify patterns that can guide decision-making processes and bolster student learning experiences [2]. While there are many studies that provide recommendations to enhance student experiences in educational settings [3,4,5], gauging student learning and its intricacies remains a challenge. Traditionally, examinations have served as a favored assessment tool to evaluate various facets of students' abilities, such as knowledge, skill, aptitude, and physical capabilities [6]. Performance in these examinations often provides insights into a student's likelihood of dropping out [7]. As such, delving deep into the analysis and prediction of examination outcomes is crucial. This study aims to consolidate insights from past meta-analyses and offer a comprehensive review of analytical and predictive methodologies explored over the past two decades.

2. Background

2.1 Examinations

Examinations are typically defined as a method to determine and evaluate progress or the current state of knowledge and learning [8]. An examination could be structured in multiple ways. The most common types of examination methods include an oral examination, written examination, or hybrid examination which involves different methods [9].

Over the past few centuries, the primary means of evaluating human knowledge have remained surprisingly consistent. Oral and written examinations are still the most used methods. Examinations are an extremely popular form of passing knowledge. This stems from the fact they are quite reliable in predicting the outcome of learning or teaching [10]. It is often most convenient and effective to test the knowledge that was administered. Although there might be cases where the real-world applications may differ from the theoretical conjunction given the simplicity, examinations continue to be a preferred choice.

They not only serve as vehicles of assessment but also as instruments of encouragement and provide validation to a learner's knowledge. This helps the learner have self-awareness in the next phase of growth for them. It encourages hard work and critical thinking and tests the learner in difficult conditions as well [11].

Examinations are critical in a learner's pursuit because they serve as an important tool for the learner to assess their knowledge in a controlled environment [10]. It helps in standardizing systems across a large country, state, or city. It serves as a benchmark and a signal to potential future educational institutions and employers [11]. One of the other important functions of examination is to distribute limited resources. Typically, the students are awarded merit-based scholarships or admissions based on multiple factors of which examinations form an important part. This creates a competitive environment and also encourages all the learners to excel in their work resulting in a higher quality of learning [12].

2.2 Data Science in Education

Data Science is the application of multiple disciplines of mathematics, science, computer algorithms, numerical methods, data models, and empirical methods that help in analyzing meaningful patterns in data [13]. It has proven to be extremely versatile across all fields, including education. Data science in education deals with the development of methods to explore data originating in an educational context and is also referred to as educational data mining [14].

In the field of education, data science can be used to improve teaching and learning outcomes. For example, it can be used to (1) analyze student performance data to identify areas where students may need additional support or to develop personalized learning plans [15], (2) evaluate the effectiveness of educational programs and interventions [16] and (3) improve the fairness and validity of educational assessments like the ACT and SAT by detect and correcting for bias in test questions or to ensure that test scores are comparable across different test administrations [17].

2.3 Predictive Technologies

Predictive technologies are a collective set of technologies that can be used to determine an outcome with a reasonable level of accuracy. This is done by collecting large sums of historical data and exploring based on the given conditions. An example of predictive technology includes weather predictions. Scientific institutes collect data over multiple years and then administer their recommendations-based models that mimic the captured data. Predictions are critical as they can assist in saving resources, improving customer satisfaction, and ultimately helping gain the trust of its users.

Common predictive models that are used include (1) K-Nearest Neighbor (K-NN): A simple yet effective algorithm used for classification and regression tasks that finds the "k" nearest data points to a given input and makes predictions based on their values. K-NN is versatile and can be applied to a wide range of problems [18]; (2) Linear Regression: A fundamental statistical method used for predicting a continuous outcome variable based on one or more input variables. It assumes a linear relationship between the inputs and the target variable, making it widely applicable in fields such as economics and finance [19]; (3) Random Forest: An ensemble learning technique that combines multiple decision trees to improve predictive accuracy and reduce overfitting. It is particularly effective for complex data sets and is widely used in data science competitions and real-world applications [20]. These models are most commonly used due to their versatile nature [21].

The accuracy of data used to train the predictive model [22], the time required to train the predictive model [23] and the computational resources available for model training and deployment [24] are some critical benchmarks in determining the success of the model. Since success is inherently based on mathematical logic, it is important to note that the majority of data science models start their journey as mathematical logic that eventually gets converted into code. For example, linear regression, a commonly used model in data science, is based on the mathematical concept of finding the line best fit for a set of data points [25]. Similarly, decision trees, another commonly used model in data science, are based on the mathematical concept of recursively partitioning the feature space to create a tree-like structure for making predictions [26].

3. Methods

The steps used to obtain and analyze the literature for the current review were: (1) identifying the scope and research questions, (2) defining the inclusion and exclusion search criteria of literature, (3) classifying and cataloging the literature sources that relate to analyzing and predicting students' performance in examinations, (4) identifying similarities, differences and patterns in the research methodologies, data science models and results of the selected studies.

3.1 Identifying Inclusion/Exclusion Criteria

In the current study, we used the following criteria to filter literature and identify the most relevant sources.

- The study must have had a published date no earlier than the year 2000.
- The study must have been available in English either in original form or by peer-reviewed translation.
- The full text of the study must have been available as open-source material and could be retrieved through Imperial College London University's digital library.
- The study must have experimental research as the basis to test the hypothesis and not provide a theoretical review of past literature only.
- The study must have specified the research methodology, data collection process, data analysis, and results.

3.2 Finding and Cataloging Sources

The keywords used in the search criteria were "analyzing" and "predicting" and "students' performance" and "examinations". The keywords were searched on "scholar.google.com" and the publication dates were set to the years 2000 and onwards. From the results that showed up from the search, 25 studies were selected based on their title. After reading through the abstract and table of contents of each study, 10 studies out of the initial 25 were selected based on their relevance to our inclusion/exclusion criteria.

3.3 Data Analysis

Once the final set of 10 papers was compiled, each of the papers was individually qualitatively analyzed to extract basic information like study design, research question, target population domain, and learning outcomes as mentioned in Table 1. The citation in IEEE was also retrieved during this step to be used in the current study's Bibliography. The papers were then collectively analyzed for any similarities, differences, or patterns. The findings are discussed in the next section.

Table 1. Summary of Selected Articles

Article	**Study Design**	**Study Question**	**Target Population**	**Domain**	**Learning Outcomes**
[27]	Empirical study	Can we predict student performance in computing courses based on programming behavior?	Undergraduate students taking introductory computing courses	Computing education	Develop a model that can predict student performance in computing courses based on their programming behavior
[28]	Empirical study	Can we predict student performance from LMS data?	Students taking blended courses using Moodle LMS	Higher education	Develop a model that can predict student performance from LMS data
[29]	Literature review	What are the different methods for predicting academic performance?	Students of all ages and levels	Education	Summarize the different methods for predicting academic performance
[30]	Empirical study	Can we predict school performance and early risk of failure from an intelligent tutoring system?	Students in grades 5-12	Education	Develop a model that can predict school performance and early risk of failure from an intelligent tutoring system
[31]	Empirical study	Can we predict students' academic performance in	Students applying to graduate business	Business education	Develop a model that can predict students' academic performance in

		graduate business programs using admission criteria?	programs		graduate business programs using admission criteria
[32]	Empirical study	Can we predict student performance by using data mining methods for classification?	Students of all ages and levels	Education	Develop a model that can predict student performance by using data mining methods for classification
[33]	Empirical study	Can we analyze student performance in distance learning with genetic algorithms and decision trees?	Students taking distance learning courses	Distance education	Develop a model that can analyze student performance in distance learning with genetic algorithms and decision trees
[34]	Empirical study	Can we predict student performance in Massive Open Online Courses using deep learning system based on learning behaviors?	Students taking Massive Open Online Courses (MOOCs)	Online Education	Develop a model that can predict student performance in Massive Open Online Courses (MOOCs) using deep learning system based on learning behaviors
[35]	Empirical study	Can we apply data mining techniques for predicting student success in English exit exam?	Students taking the English exit exam	Education	Apply data mining techniques to predict student success in English exit exam
[36]	Empirical study	Can we predict performance in an introductory programming course by logging and analyzing student programming behavior?	Undergraduate students taking introductory programming courses	Computing education	Develop a model that can predict performance in an introductory programming course by logging and analyzing student programming behavior

4. Results

The final data set is comprised of a total of 10 papers that meet our criteria. The literature revealed interesting patterns and insights about (1) typical researchers' motivation to explore this topic, (2) common research context (location, teaching format, courses, etc.) used for experiments, (3) typical data collection methods used, (4) data parameters commonly used for data analysis, (5) data science methods employed by researchers, (6) trends in conclusions of the research and (7) agreement/disagreement in the scope of future work in this field. These are discussed in detail below.

4.1 Trends in Motivation for Research

The top motivations amongst researchers to analyze and predict students' performance in examinations were to: (1) identify students at risk of failing or dropping out and (2) guide admission decisions based on the prediction of which students will do well in school. If the students at risk of failing examinations or dropping out are identified earlier, targeted interventions can be implemented by institutes to guide them to success.

Analyzing and identifying which students will perform well in an institute is beneficial for the admissions department, so they can minimize students who fail and drop out and maximize students who will excel, succeed, and elevate the institute's reputation. The data analytics can also reveal or verify if the existing university admissions criteria are in line with university standards. Other common motivations mentioned in the papers are as follows:

- *Identifying students who should be screened for learning disabilities*. Haridas et al. [30] in their paper talk about how many primary students in India are not assessed for reading difficulties. Even though many factors such as lack of awareness, training regarding learning difficulties, shortage of special education teachers and concerns about labeling the child contribute to this phenomenon [37, 38], the cost of clinical evaluation remains a critical one. By correctly identifying students who might have learning disabilities, we can reduce the number of clinical evaluations that need to be conducted and hence the associated cost.
- *Guiding decisions on how to organize the educational institutes' marketing campaigns and approach promising potential students.* Kabakchieva [32] in their paper talks about how university management can analyze the profile of admitted students for their specific characteristics and use to it create targeted marketing campaigns for promising students.
- *Guiding decisions for providing financial aid to students.* By analyzing how students perform on their admission tests and predicting how well they are likely to perform throughout their course, educational institutes can determine the best candidates for financial aid [29].
- *Improving instructional methods.* The students' performance prediction can not only help identify weak students and provide tailored support to them but also help instructors improve course material, teaching methods, or technology [27]. They can identify which methods of instruction best translate into student learning and content retention and also identify which course topics most students have trouble understanding. The instructors can then use this information to focus more on those topics by including interactive

activities or simplifying the concepts.

- *Existing models of analyzing student performance are not applicable in software programming domain.* According to Watson, Li and Godwin [36], the limitation of studies to date is that they use lengthy tests that often produce inconsistent results when ported to programming courses. As the number of enrollments is increasing in programming courses, the use of these tests to gather data can take an instructor significant time to process. By the time a helpful prediction is indicated, it might be too late for a student to withdraw from the course or for an instructor to intervene and provide tailored support. The student traits such as cognitive, psychological, behavioral, or demographic may be helpful indicators of performance in other domains, but they take time to assess and are not directly related to the general programming behavior of a student. Therefore, new predictors of performance need to be explored in the programming domain that are not based upon indirect criteria but are rather based upon criteria which can be automatically measured and relate best to the programming behavior of students to increase the accuracy and speed of the feedback process.
- *Lack of assessment technologies in MOOCs (Massive Open Online Courses).* According to Lee et al. [34], in MOOCs, students cannot be monitored in real-time, there is scope for cheating in exams, and learners are extremely diverse in terms of prior knowledge and absorb different content. Therefore, the existing predictive models provide inconsistent results. Recent research reports high dropout and low completion rates for MOOCs [42], and therefore, new solutions need to be explored.

4.2 Research Context

The research context for the scope of this paper is defined by fundamental attributes of the target population (1) physical location, (2) age, (3) sample size, (4) primary instruction format, and (5) assessed subjects. The literature review reveals that there is a global interest among researchers in analyzing and predicting students' performance in examinations. In the ten papers reviewed, the research is focused on regions - the United States of America, the United Kingdom, Argentina, India, Greece, and Bulgaria. There is a mixture of both urban and rural locations as well. For example – the study by Haridas et al. [30] focuses on 2123 students at 5 rural Indian schools who used the intelligent tutoring system, AmritaITS. Most of the studies are focused on undergraduate students while other specific categories explored include K-12, graduate business schools, and open universities (with both undergraduate and graduate courses).

The sample size of students in the study varied greatly as well. Kabakchieva [32] studied a large set of 10330 students at a Bulgarian University while a UK-based study by Watson, Li and Godwin [36] only analyzed 45 students in a programming course. The low number of students in the analysis was attributed to the fact that only 45 students provided permission to rack their learning activity.

There is diversity in instructional formats as well. The formats reviewed include in-person teaching, online teaching, ITS (Intelligent Tutoring Systems), hybrid teaching and MOOCs. An interesting insight gained from the review was that only a couple of studies focused on the general prediction of students' performance, while most of the studies focused on a specific subject or domain. For example – the study by Puarungroj et al. [35] only focuses on performance in English exit exams, the study by Watson, Li, and Godwin [36] only relates to programming courses and Haridas et al. [30] talks specifically about predicting performance in English and Mathematics in their research. The trend of conducting domain-specific prediction might relate to the poor portability of prediction models. Conijn et al. [28] found in their study that there does not exist a single best way to predict student performance across a diverse set of courses. This is also in line with Carter, Hundhausen, and Adesope’s [27] argument about developing a special prediction model for programming courses due to different relevant data parameters (detailed discussion in the conclusions section).

4.3 Common Data Collection Methods

The primary methods of data collection used in the studies include (1) surveys, (2) assessments, and (3) tracked activity in ITS.

Surveys have been used to gather student information such as gender, birth year, birthplace, living place, type of previous education, profile, place of previous education, and the total score from previous education along with educational institutes-maintained records on admittance year, educational major or specialty, the current semester, along with assessment scores such as admission tests, homework, quizzes, and examinations. The assessments could be in a pen-paper format or digital, multiple-choice or textual, in-person or virtual.

A modern data collection method discussed in some of the studies was tracking online activity in an Intelligent Tutoring System (ITS). An ITS simulates a one-on-one interaction with an educator and provides feedback and instructions to students without the intervention of human educators [39]. ITS outperformed other comparable modes of instruction during meta-analysis of evaluative studies and has become a popular choice of instruction due to the rise in demand for effective online learning methods [40, 41]. AmritITS, used in the study by Haridas et al. [30], was one of the ITS systems utilized for data collection. Another interesting approach to data collection was used by Carter, Hundhausen, and Adesope [27] in their study where they used OSBIDE, a plugin for Microsoft® Visual Studio® to track students' programming activities.

4.4 Parameters Used for Data Analytics

Data variables and features used for research showed maximum variance in the research papers. Even though some of the features like prior examination scores were universally used, most of the data variables were very specific to the study. There was also a huge difference in the number of parameters analyzed for predicting students’ performance. While one study used about 22 variables, another study used more than 100.

Lee et al. [34] developed the maximum number of features. In their study, they recorded and analyzed each student’s answer to all exercise questions and extracted eight features such as exercise type (single, multiple, fill in the blanks), number of correct answers, number of attempts, time to complete the exercise, if a student watched a related video before answering and if a student watched a related video after answering correctly. They also extracted course information: duration, number of students, number of videos, number of exercises, number of quizzes, quizzes interval time, fees and developed video-watching features such as the proportion of videos finished, time spent watching videos, number of days per week spent on learning amongst others.

Puarungroj et al. [35] defined features as sex, faculty, blood type, entry GPA, and English placement test score to predict student performance in the English exit exam. Hoefer and Gould [31] analyzed GMAT score, birth date, sex, quant score, verbal score, TOEFL score, and the tier of school of bachelor education to evaluate the criteria for admission into graduate business programs. Conijn et al. [28] pre-processed raw Moodle (a popular Learning Management System (LMS)) log data using R to create predictor variables such as total number of clicks, number of online sessions attended, average time spent per session, number of discussion post views, the largest period of inactivity, irregularity of study time, irregularity of study interval, etc. along with assessment data for midterms, quizzes, reports, assignments, and homework.

4.5 Typical Data Analytics Models Employed

The traditional linear regression model was the most popular choice for modeling. Other common models included simple neural networks, deep neural networks, and multiple linear regression (MLR). Haridas et al. [3-] utilized mixed-effects logistic regression models. to predict at-risk students failing the final exam.

The metrics for model evaluation: (1) R Square / Adjusted R Square, (2) Mean Square Error (MSE) / Root Mean Square Error (RMSE) and (3) Mean Absolute Error (MAE) were consistently used across all studies to assess the accuracy of the fit of the models. Two of the studies used 10-fold and 20x5-fold cross-validation techniques to verify the robustness of the models [28, 30].

A high variance was observed in selecting the classification algorithms in the studies. Kalles and Pierrakeas [33] argued that genetic algorithm-based induction of decision trees is a superior classification tool to analyze student performance in student learning while Kabakchieva [32] experimented and compared J48, jRip, kNN 100, kNN 250, Classifier, NaiveBayes and BayesNet classification algorithms on their data set to reveal a high potential of data mining applications for university management.

An interesting measure of prediction was discussed– the Watwin Score or Watwin Algorithm – designed especially for programming courses [27, 36]. It focuses exclusively on differences between successive code compilation attempts. Metric associates improved learning outcomes with an ability to quickly remove compilation errors from a program.

A mention of the Cross-Industry Standard Process for Data Mining (CRISP-DM) model was also found [32]. The CRISP-DM is a cyclic approach with six phases and is popularly chosen as a research approach because of its non-propriety nature, free availability, and application-neutral standard for data mining projects. The model includes several internal feedback loops between the phases to ensure the achievement of consistent and reliable results.

4.6 Patterns in Study Conclusions

The studies in the current literature review tested different aspects of analyzing and predicting students' performance in examinations. Some of the common conclusions the studies arrived at are discussed below.

- *Portability of prediction models.* Studies found that there does not exist a single best way to predict student performance across a diverse set of courses [28]. Therefore, it is often required to develop a special prediction model for programming courses due to different relevant data parameters [27].
- *Admission score as a significant predictor of student performance in the educational institute*. Studies found that university admission scores like placement tests and GMAT are among the factors influencing the classification process the most and therefore, an important determinant of students' academic performance [32, 35, 41].
- *Need to consider "humanistic" attributes*. Studies revealed the need for a holistic approach to analyzing data. For example - Hoefer and Gould [31] showed how using utilized qualitative indicators in their neural network model produced useful insights like students graduating from tier 1 undergraduate schools tended to have better academic performance in graduate business school than graduates of tier 2 to 4 schools. Similarly, the study by Carter, Hundhausen, and Adesope [27] showed that their holistic model - the Normalized Programming State Model (NPSM) performed much better at predicting student performance in programming courses than the Error Quotient and Watwin Score.
- *Focus on intelligent systems.* All studies reveal the importance of incorporating intelligent systems for analyzing performance. For example - Kalles and Pierrakeas [33] discussed in their study how genetic algorithm-based induction of decision trees could be used as an "early warning system". This would notify educators well in advance of the final exam about the expected students' performance and start a remedial intervention accordingly. Lee et al. [34] also developed a novel AI system that could remedy the present-day inability of MOOCs to evaluate student performance. The AI system evaluates the learning behaviors of the learners and predicts the proportion and types of questions students will answer correctly, thus indicating the precise concept the learner is struggling with or excelling at.

4.7 Consensus on Scope of Future Work

A common theme that the studies discussed for future work was scaling the predictive models to different courses and learning environments. For example - Carter, Hundhausen, and Adesope, [27] discussed applying the NPSM to different programming languages, environments, and computing courses as the next logical step in research.

Another theme that emerged was further validating the proposed solutions. Lee et al. [34] proposed that future studies should consider applying the proposed AI-based evaluation system to other MOOCs to validate its effectiveness using larger datasets. In this case, the proposal is not to port the suggested model to a different learning environment but to validate it in the same environment but on a larger data set. Similarly, Watson, Li, and Godwin [36] suggested that future work should aim to further validate the presented dynamic algorithm, Watwin, by applying it to a different set of data gathered from an independent sample of students.

Finally, many of the studies discussed developing an intelligent tool based on their research in the scope of future work. The proposal is to integrate different aspects of predictions and develop a piece of software that can be used to dynamically analyze students' performance and provide feedback. This software could be as simple as an analytics dashboard or as complex as an AI-based ITS.

5. Conclusion

The goal of the current was to review and consolidate the most up-to-date literature on analyzing and predicting students' performance in examinations. For this, we performed a literature review using ten papers that were selected based on the specified inclusion and exclusion criteria. After qualitatively analyzing the papers, it appears that the experimental research in this field is still in the nascent stages. Even though the researchers have been developing and testing models for the past three decades, there are only a few studies that have been able to apply and verify their research on a larger scale.

All the studies talk about the huge potential for developing intelligent systems, but this potential is not vastly explored. The COVID-19 pandemic brought the limelight to virtual teaching systems and thus the need for learning management systems (LMS) and intelligent tutoring systems (ITS) surged. There are some big market players like Brightspace, Moodle, and Blackboard in the LMS domain, however, there are only beta-phase products like AmritITS and ElectronixTutor in the ITS domain.

Similarly, the importance of utilizing qualitative aspects of student learning behavior is broadly discussed, however, only a few studies use humanistic attributes in their models. We believe that this aspect of educational data mining (EDM) needs to be further explored. More ways to collect, quantify, and model qualitative data need to be developed and validated.

Another observation from the studies was that there was a specific set of subjects the researchers were interested in predicting student performance about such as computer programming, STEM, and English. This is not surprising since the education community has seen an increased focus on software development and STEM education in the past decade. However, it will be interesting to see how the current prediction models perform in non-STEM subjects such as business management, finance, sociology, psychology, hospitality management, nursing, and fashion. Business management graduates and finance graduates manage important aspects of organizations, and their university admission and company interview process are often vague, stemming from the fact that their performance cannot be easily accessed through traditional methods and the knowledge is considered subjective. However, with the new prediction models that consider qualitative aspects of student learning as well, new parameters can be developed that can help admissions as well as organizations to hire the best candidates.

6. Discussion

This literature review emphasizes the growing emphasis on leveraging data science and predictive technologies to enhance educational outcomes. The synthesis of research findings reveals the multidimensional challenges and opportunities in predicting students' performance in examinations.

One intriguing aspect is the need to strike a balance between algorithmic predictions and humanistic considerations. While models can effectively analyze historical data and offer predictive insights, human-centric attributes such as student background, qualitative indicators, and learning behaviors remain crucial. This aligns with the contemporary trend of personalized education, where data analytics informs decisions while acknowledging the uniqueness of each student.

The diverse methods for data collection and analysis also reflect the evolving landscape of educational technology. The utilization of Intelligent Tutoring Systems, online platforms, and programming tools showcase the integration of data-driven insights into the learning experience. However, as technology advances, ethical concerns regarding data privacy and algorithmic bias need careful consideration to ensure equitable educational opportunities for all students.

The identified research gaps, such as the need for cross-course validation and intelligent tool development, open avenues for future inquiry. As predictive models mature, their successful deployment depends on seamless integration into educational practices. The proposed intelligent tools hold promises for providing actionable insights to educators, administrators, and students in real time. However, their efficacy and usability warrant rigorous testing and iterative refinement.

In conclusion, the intersection of data science, education, and predictive technologies offers a transformative potential to optimize student learning and success. By harnessing the power of data analytics, researchers and practitioners can drive evidence-based interventions, empower educational institutions, and ultimately enhance the educational journey for students worldwide.

References

[1] NCES, 2013. National Centers for Education Statistics, U.S. Department of Education.

[2] S. Aljawarneh and J. A. Lara, "Data science for analyzing and improving educational processes," *Journal of Computing in Higher Education*, vol. 33, pp. 545–550, 2021.

[3] R. Ammigan and E. Jones, "Improving the Student Experience: Learning From a Comparative Study of International Student Satisfaction," *Journal of Studies in International Education*, 22(4), 283-301, 2018.

[4] C. Alvarado, M. Minnes, and L. Porter, "Micro-classes: A structure for improving student experience in large classes," In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, 21—26, 2017.

[5] S. Aggrawal and A. J. Magana, "Undergraduate student experience with research facilitated by project management and self-regulated learning processes," in 2023 ASEE Annual Conference & Exposition, 2023

[6] P. Rawlusyk, "Assessment in higher education and student learning." *Journal of Instructional Pedagogies*, vol. 21, 2018.

[7] C. B. Davison and G. Dustova, "A quantitative assessment of student performance and examination format." *Journal of Instructional Pedagogies*, vol. 18, 2017.

[8] Merriam-Webster.com, 2019. Definition of EXAMINATION. [online] Available at: https://www.merriam-webster.com/dictionary/examination.

[9] A. Ghaicha, "Theoretical framework for educational assessment: A synoptic review." *Journal of Education and Practice*, vol. 7, no. 24, pp. 212–231, 2016.

[10] C. Gipps, *Beyond Testing (Classic Edition): Towards a theory of educational assessment*. Routledge, 2011.

[11] P. Black and D. Wiliam, "Assessment and classroom learning," *Assessment in Education: principles, policy & practice*, vol. 5, no. 1, pp. 7–74, 1998.

[12] W. Harlen and M. James, "Assessment and learning: differences and relationships between formative and summative assessment," *Assessment in Education: Principles, policy & practice*, vol. 4, no. 3, pp. 365–379, 1997.

[13] National Academies of Sciences, Engineering, 2020. Roundtable on Data Science Postsecondary Education: A Compilation of Meeting Highlights. Publications listing National Academy of Sciences National Academy of Engineering Institute of Medicine National Research Council. Retrieved from https://par.nsf.gov/biblio/10295204.

[14] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, vol. 40, no. 6, pp. 601–618, 2010.

[15] M. Bienkowski, M. Feng, and B. Means, "Enhancing teaching and learning through educational data mining and learning analytics: An issue brief." *Office of Educational Technology, US Department of Education*, 2012.

[16] R. S. Baker, K. Yacef, *et al.*, "The state of educational data mining in 2009: A review and future visions," *Journal of educational data mining*, vol. 1, no. 1, pp. 3–17, 2009.

[17] Zwick, R. (2017). Fairness in College Admissions: Assessment of Student Attributes. Educational Measurement: Issues and Practice, 36(1), 11-20.

[18] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[19] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to linear regression analysis. John Wiley & Sons, 2021.

[20] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
[21] Bush, T., 2020. Predictive Analysis: Definition, Tools, and Examples. [online] pestleanalysis.com. Available at: https://pestleanalysis.com/predictive-analysis/.
[22] F. Provost and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. ” O’Reilly Media, Inc.”, 2013.
[23] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd International Conference on knowledge discovery and Data Mining*, 2016, pp. 785–794.
[24] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
[25] G. James, D. Witten, T. Hastie, R. Tibshirani, *et al.*, *An introduction to statistical learning*. Springer, 2013, vol. 112.
[26] L. Breiman, J. Friedman, R. Olshen, and C. Stone, “Classification and regression trees–crc press,” *Boca Raton, Florida*, 1984.
[27] A. S. Carter, C. D. Hundhausen, and O. Adesope, “The normalized programming state model: Predicting student performance in computing courses based on programming behavior,” in *Proceedings of the Eleventh Annual International Conference on international computing education research*, 2015, pp. 141–150.
[28] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, “Predicting student performance from lms data: A comparison of 17 blended courses using moodle lms,” *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 17–29, 2016.
[29] M. Gallacher, “Predicting academic performance,” 2005.
[30] M. Haridas, G. Gutjahr, R. Raman, R. Ramaraju, and P. Nedungadi, “Predicting school performance and early risk of failure from an intelligent tutoring system,” *Education and Information Technologies*, vol. 25, pp. 3995–4013, 2020.
[31] P. Hoefer and J. Gould, “Assessment of admission criteria for predicting students’ academic performance in graduate business programs,” *Journal of Education for Business*, vol. 75, no. 4, pp. 225–229, 2000.
[32] D. Kabakchieva, “Predicting student performance by using data mining methods for classification,” *Cybernetics and information technologies*, vol. 13, no. 1, pp. 61–72, 2013.
[33] D. Kalles and C. Pierrakeas, “Analyzing student performance in distance learning with genetic algorithms and decision trees,” *Applied Artificial Intelligence*, vol. 20, no. 8, pp. 655–674, 2006.
[34] C.-A. Lee, J.-W. Tzeng, N.-F. Huang, and Y.-S. Su, “Prediction of student performance in massive open online courses using deep learning system based on learning behaviors,” *Educational Technology & Society*, vol. 24, no. 3, pp. 130–146, 2021.
[35] W. Puarungroj, N. Boonsirisumpun, P. Pongpatrakant, and S. Phromkhot, “Application of data mining techniques for predicting student success in English exit exam,” in *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, 2018, pp. 1–6.
[36] C. Watson, F. W. Li, and J. L. Godwin, “Predicting performance in an introductory programming course by logging and analyzing student programming behavior,” in *2013 IEEE 13th International Conference on Advanced Learning Technologies*. IEEE, 2013, pp. 319–323.
[37] R. Kamala and E. Ramganesh, “Difficulties in identifying the dyslexics in multilingual context,” *International Journal of Humanities and Social Science Invention*, vol. 4, no. 1,

pp. 18–22, 2015.
[38] A. Kohli, S. Sharma, and S. K. Padhy, “Specific learning disabilities: Issues that remain unanswered,” pp. 399–405, 2018.
[39] K. VanLehn, “The behavior of tutoring systems,” International *journal of artificial intelligence in education*, vol. 16, no. 3, pp. 227–265, 2006.
[40] J. A. Kulik and J. Fletcher, “Effectiveness of intelligent tutoring systems: a meta-analytic review,” *Review of educational research*, vol. 86, no. 1, pp. 42–78, 2016.
[41] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, “Intelligent tutoring systems and learning outcomes: A meta-analysis.” *Journal of educational psychology*, vol. 106, no. 4, p. 901, 2014.
[42] Y. Sun, L. Ni, Y. Zhao, X.-L. Shen, and N. Wang, “Understanding students’ engagement in moocs: An integration of self-determination theory and theory of relationship quality,” *British Journal of Educational Technology*, vol. 50, no. 6, pp. 3156–3174, 2019.