Mini Review

The Impact of Statistics and Probability on Educational Artificial Intelligence

Adriana Rodríguez Rosales*

School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey 64849, Mexico

https://www.mathematicsgroup.com

Check for updates

Abstract

Artificial intelligence has transformed e-learning by enabling personalized and efficient teaching. This manuscript analyzes the importance of statistics and probability in educational AI. Statistical methodologies improve decision-making, personalize learning, and optimize educational outcomes. Challenges such as data privacy and ethics are addressed. Case studies demonstrate the practical applications of AI in diverse educational contexts. Future directions suggest a need for robust research to further understand and implement AI-driven educational strategies. The findings underscore the critical role of data-driven approaches in shaping the future of education.

Statistics and probability are not only foundational to the development of AI but also essential for analyzing vast amounts of educational data. They allow for the creation of predictive models that can identify student needs and adapt instructional methods accordingly. This adaptability enhances the learning experience by providing targeted support and resources to students, thereby improving their academic performance.

Ethical considerations are fundamental when using AI to handle educational data. Protecting student data with privacy and security is crucial to maintaining trust in AI applications. This manuscript examines how educators and policymakers can collaborate to create guidelines that safeguard student information while utilizing data to enhance education.

Integrating statistics and probability into educational AI significantly impacts and improves e-learning. Educators can enhance learning by employing data-driven strategies that provide personalized and effective teaching. This approach not only benefits individual learners but also contributes to the overall advancement of educational practices. Embracing these data-driven methodologies is essential for the continued evolution of teaching and learning in the digital age.

Introduction

Currently, the use of information and communication technologies across various fields is continually evolving. One of the areas where technology has become crucial is in education, giving rise to what we know as e-learning. The growing utilization of e-learning has been enabled by Artificial Intelligence (AI), which has evolved with various features applicable to these technologies, such as data mining in e-learning (Shute & Glaser, 2010). This allows educators from various institutions to obtain indicators of student behavior and learning, with the aim of enhancing the teaching-learning processes through e-learning (Renshaw, 2012).

Contextualizing the topic

Contrasting this, we see the work of Shute and Glaser, who create educational situations where the student assumes the role of a decision-maker, reflecting their conceptual model of the problem. This stance is supported by Renshaw, who advocates for realistic trial contexts particularly relevant in vocational training environments. To optimize decision-making, students must evaluate differences and similarities, skills that are honed by enhancing their statistical and probabilistic capabilities, where historically computers have played a significant role as facilitators. Acquiring competencies in mathematics is a critical issue in both basic and secondary education. According to the SIMCE 2005 study by the Ministry

of Education of Chile, a low percentage of students achieve adequate and significant performance in Mathematics, notably in 12th grade, the lowest among the five disciplines. These results indicate that a basic component of the average Chilean student's mathematical competence involves handling arithmetic and algebraic calculation techniques, with serious deficiencies in other components of their competence. Renshaw points out that when dealing with real-world problems, the solution will address continuous variables with equal probability, representing an unfounded choice if applied to a real-world problem.

Research significance and questions

This research is significant as it addresses the integration of statistics and probability within the domain of educational AI, a critical aspect for enhancing data-driven decision-making in education. The study aims to answer the following research questions:

1. How can statistical methods improve the personalization of learning experiences in AI-driven educational systems?

2. What are the ethical and privacy implications of using AI in education?

3. How can educators leverage AI to optimize pedagogical outcomes while ensuring data security and privacy?

Foundations of Statistics and Probability in Artificial Intelligence Among the explanations on learning, Genaro Guerra provides one of the most powerful explanatory approaches when he defines learning as a change in an individual's behavior repertoire that is relatively permanent and occurs because of experience or practice: a change occurs in knowledge, beliefs, motivations, attitudes, emotions, and physical ability. Undoubtedly, changes in the behavior repertoire and being relatively permanent are the most notable definitions offered, both in reference to the concept itself and the keys to its definition, as they encapsulate a fundamental idea of the concept of learning. The realm of Educational Artificial Intelligence (EAI) is fundamentally grounded in the disciplines of Statistics (S) and Probability (P). Of course, these are not the only foundational disciplines for the advancement of EAI, as contributions from related fields in computing and intelligent systems must also be considered. However, methodologies for evaluation (based on error calculation), categorization, and decision-making (especially personalization) based on models related to various variables that can intervene in the statistical versions of the students themselves are vital for incorporating Machine Learning (ML) techniques in EAI systems close to the rule-based performance.

Basic concepts of statistics and probability

The importance of teaching statistics and probability is evident to appreciate both the specific context, as there are probabilistic generalizations inherent to different disciplines, and the overall relevance in comprehensive education. In the workplace, statistics become a highly relevant tool for a wide range of business and corporate issues. The subject of statistics and probability is perceived as challenging by a high percentage

of students according to the differentiation schemes of HUZ (1985), within the mathematical sciences, sharing these with number theory, algebra, and calculus. Probability is the study of randomness, of uncertainty. In a broader sense, the theory of probability means the abstraction of real life but is used to model and understand it, used to create a model. Probability surrounds us. According to Laplace (1758 - 1827), "The theory of probability is nothing more than common sense reduced to calculation." With this, Laplace asserts that probability is a science based on pure reason. While it is clear that many issues concerning uncertain situations are known to us by habit, and that probability in many situations is sufficient with intuition, it is always important to record how these appreciations or conclusions were obtained.

Practical applications in educational artificial intelligence

These differences can negatively impact the management of interactive decision-making processes and, consequently, pedagogical aspects, unless they are identified and managed correctly. Recommendations emerging from this review include promoting large-scale research with EAI to address diverse and explicit educational aspects, as well as to test interventions in different educational contexts and levels. Balancing basic and engineering research to advance the science and technology of EAI is crucial. Consistency among different areas and branches of EAI, and between academic and programmed interventions, is essential. Limiting ad hoc studies that potentially affect the reliability of the tests and directing studies based on evidence is necessary. We have implemented EAI applications in various educational contexts, such as highlighted interventions on early warning systems and pedagogical reports. These student-focused interventions have proven effective in the context of secondary education in Egypt, optimizing the accuracy of the early warning system on school dropouts and refining personalized pedagogical reports at the level of suggestions. However, there are still open challenges concerning the identification of moderating differences and the design of interactive inference with the user. For instance, responses may differ for different student activity analysis systems (ALS) and/or evaluation systems designed in specific educational contexts or aimed at certain aspects of the educational process.

Personalization of learning

As previously cited, leveraging the versatility of technology and adapting content and teaching and learning processes to the peculiarities of each student makes it possible to personalize learning (instructional inspiration). This fact is particularly important since the diversification of the student body makes it unfeasible for instructional action to be directed at the peculiarities of each individual unless it can be personalized for each case. Therefore, it is necessary to emphasize that personalization must be understood in a strictly educational sense. Thus, personalizing teaching becomes a fundamental facet of the "plan" or "broad guideline," while the "non-personalizable" refers to the realm of that planning that does not affect the cognitive and personal aspects of the students (for example, the tone used when saying "good morning" to the class). The teaching that takes place in a classroom and

its corresponding design is carried out by the teacher, so that, to a greater or lesser extent, the material and the sequence of learning are the same for all students. However, educational technology, and specifically artificial intelligence, poses the design of teaching systems in which the first condition, i.e., the connection between the knowledge to be taught and to be learned, is unique (corresponds to the same concept). However, the remaining conditions, particularly those of the students (to order/sequence/dose the material in accordance with their knowledge, capability, interest, etc.), may be different and thus face the as-yet indescribable challenge of designing a learning environment finely personalized enough for its recipients.

Challenges and future of integrating statistics and probability in educational AI

In this way, the emerging faculty of DDM (Data-Driven Models) must be preceded by a Data-Literate Faculty (data scientists) highly specialized in the statistical and probabilistic aspects of data. Undoubtedly, when a high school student is presented with the problem of the flight time of a projectile in a parabolic throw, or when an economics science teacher chooses to exemplify the utility of statistics in the descriptive, interpretative, and explanatory part of an economic principle, a possible scenario of predictions about the production of a good or service; or, the mathematics teacher, who presents the accounts with the profits from the operation of a bank after providing many quantitative elements that define the problem, to decide the formulas that will describe the profits according to the initial amount deposited, making decisions and modifications after describing the environment, will enable personalized speeches and empower students in their learning processes. By incorporating statistics and probability, we would not only cover a quantitative aspect by posing the situation in this way but also engage in a qualitative inquiry about the problems. It can reduce the likelihood of opting for a model that does not meet the desired generalization conditions. It can even explain the internal processes of the models and their necessity. This same idea can be applied to the division into training, validation, and test sets, where the aim is to replicate the later performance of the model with a dataset different from a model trained with a specific set. Undoubtedly, skillful handling of statistical concepts can be crucial in making decisions related to the models.

Ethics and privacy

Several aspects at the ethical (decision-making related to data treatment) and legal (set of legal norms that regulate human behavior in a social environment, establishing the application of sanctions) levels need to be considered. One of the keys lies in the concept of privacy, which can be understood from two angles: public or social, referring to the existence of norms derived from values and social customs that set the boundaries or limits of sharing and protecting information among people; and individual, which would necessarily combine with the previous basis, but referring to the individual's ability to decide the extent to which they wish to share their information (in relation to the conception of the self, as part of the keys to private behavior, or before the social environment). Some EAI applications, especially related to intelligent learning systems capable of presenting learning materials or adapting teaching methods to the learning abilities/predictions of students, feature a high level of personalization that leads to their use as tutorial systems, individually based and apparently intelligent. However, the power of prediction comes from processing and combining different types of data related to the activities and characteristics of the students (social support data, personal information, multimedia, resources, evaluation, communication, collaboration support tools, learning evolution, and interactions). The use of these data can pose a threat to the privacy of students, requiring significant protective measures. Moreover, the processing of this data can lead to the violation of equally important legal aspects such as the protection of personal data or the right that each entity (centers, associations, or departments) may have to exploit or disclose knowledge potentially generated by student-generated information.

Case studies and results

Testoni, Capodanno, and Canova [1] conducted a study on the behavior of 30 Italian students aged 9 to 12 years while working in an educational environment rigorously enriched with AI and AI-Ed, focusing the study on three individuals per week who worked in class, and asking them to reflect on their perceptions and emotions. The qualitative data collected (records, photos, field journals) were studied using Tessa's theory of emotions, Atkinson's pragmatic theory of mind, and Karpman's drama triangle model as tools to interpret and analyze the narrative of the qualitative data. The results showed that the students perceived the technologies as sentient, animated, and with a particular selective or instructive role. A case study constitutes a relevant strategy for social sciences, humanities, administration, and, in general, for all research aimed at penetrating the knowledge of particular phenomena. A case study is an empirical investigation that examines a phenomenon (events, incidents, actions, or organizational decisions) based on data collection. Among its benefits are the generalization of results from the theory, providing evidence to increase the veracity of a theory, and offering the opportunity to examine complex phenomena in natural contexts. The application of AI technologies in educational environments has been extensive. Some studies, using adaptive environments such as WISE (Webs of Inquiry and Explorations), knowledge-based tutoring like AutoTutor, quiz-based like Sheikh, AffectiveTutor, SCOR (Student Content Objectives Recommender), and authoring tools like WELSA, CAME, and RHETORIS, among others, have highlighted the benefits of AI in enhancing students' abilities to handle the information provided by the system. However, few studies have delved into the impact of statistics and probability on educational artificial intelligence [2-4].

Data analysis on learning platforms

In this way, the experience can be improved, and learning can be enhanced, especially in the distance education modality using LMS. Intelligent systems, such as Artificial Intelligence Systems (AIS), require the analysis of large amounts of data coming from various sources. Similarly, it is necessary to store this information in a database that allows the flow and processing of information. A concrete example of an LMS platform is Moodle or in the case of CMS, WordPress. Moodle collects data corresponding to the courses to then store such

information through the Log plugin. Moodle allows the teacher to view the course progress (through reports) that includes the number of attempts the student has made, grades, and even the time they spent on the platform or between tasks - for this, the teacher must configure the course tracking system. Learning platforms, commonly known as LMS (Learning Management Systems), have become an important resource in the teaching-learning process in various educational spaces. A learning platform can process large volumes of data, such as the tests that students solve, the searches they perform on a specific topic, the number of clicks they make, or the amount of time they spend on a given activity, among others. The management of large volumes of data has enabled the identification of elements linked to student achievements, both descriptively and predictively. Data analysis allows for better profiling of both individual students and groups of students with similar characteristics. This leads to differentiated treatments of the information provided to them, depending on their characteristics, or to a better design of learning activities and the optimization of educational resources and time used.

Conclusions and recommendations

More attractive training programs are needed, real in terms of the unique problem presented by the capture, selection, and processing of information, and close to the disciplinary or professional reality of the student. It will be important to integrate, or at least compare, training in different tools. Only in this way will it be possible for the student, once graduated, to have a favorable opinion towards Statistics, primarily based on issues of educational order. They understand and value the depth of reflection it provides, the potential in information processing it offers, and the importance of decision-making and problem-solving. Someone who can fit it into their discipline and professional reality, and with whom they may not always connect but at least will not disagree in terms of its certainly instructive character. Ultimately, someone who enjoys solving problems using statistical techniques and who does not mind accepting a challenge in exchange for a respectable question and, why not, another interesting doubt. The analysis of statistics and probability in the field of EAI allows us to conclude that they are increasingly accepted and required to understand what is happening and why with the teaching-learning process or with the use of an intelligent tutorial system. This separation has been, and largely continues to be, the result of students' unfamiliarity with the subject. Therefore, students are unaware that the objective science that the English taught throughout their previous educational stage was 90% incorrect, and members of the Humanities, Social Sciences, or Education Sciences disciplines do little to help them overcome this ordeal. On the contrary, they are dedicated to denying the role of statistics in the research of their disciplines and living anchored in outdated and nearly irreproducible positions.

Summary of advances and potentialities

From what has been analyzed, the possibility offered by current ICT developments, especially AI, as agents capable of providing training to individuals with much higher levels of personalization according to each one's characteristics,

whether educational or related to the structural and functional aspects of the brain, is commented upon. Attention can be focused on aspects of the new developments of constructivist, connectivist and transdisciplinary educational applications supported by the use of statistics and probability, which would allow teachers and students to influence the design of artificial intelligence. To this end, a more critical and reflective attitude should be fostered in the design of evaluations, which enrich and favor this attitude, since detecting, for example, a good degree of abduction requires a much more enriching and complex conceptual framework than simply the use of a particular technique. In summary, statistics and probability are essential for different dimensions of contemporary life. Some aspects of these links are highlighted, excluding those that pertain to mathematics and statistics education, without forgetting that in addition to these links, they also contribute directly and indirectly to other fields of application. In particular, the importance of teaching statistics and probability in this context of contemporary life is emphasized.

Acknowledgment

References

1. Testoni E, Capodanno S, Canova L. The impact of AI in educational settings: A case study in Italian primary schools. J Educ Technol. 2019;14(3):45-60.

2. Baker RS, Siemens G. Big data and education. Teachers College, Columbia University; 2019.

3. Holmes W, Bialik M, Fadel C. Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign; 2019. Available from: https://discovery.ucl.ac.uk/id/eprint/10139722/

4. Santos OC. Machine learning in education: A comprehensive overview. Springer; 2016.