Deep Science Publishing
https://doi.org/10.70593/978-93-49307-42-1_5

Chapter 5: Applying big data analytics to enhance curriculum development, administrative efficiency, and sustainable resource planning in education

5.1 Introduction

With the emergence of the knowledge economy, which is fast becoming intertwined with globalization and a growing technological revolution, it is widely acknowledged that the knowledge and skill levels of the population are of vital importance. Policymakers in many countries are keen to increase participation in secondary and higher education and training. Every child is supposed to be in school; many are encouraged to pursue courses to become more science literate and numerate, and there is a broad area of demand for a general population that has basic knowledge and understanding of higher-order thinking and processes.

But how does the community know if the educational goals are being achieved? What evidence is there to inform policy and practice? The search is on for measures, assessments, and indicators. Knowing if treatment works is fundamental to both health and educational sectors. However, we rarely enjoy the benefit of scientific random clinical trials in the educational subset. There are indicators and data. These are occasionally used judiciously, and usually either for descriptive purposes or for making an investment case. Investment cases can be based on prima facie results in league tables and a host of surrogate indicators.

5.1.1. Overview of Big Data's Impact on Educational Practices

Big data and analytics have already transformed healthcare, the military, business, and government. The capability of learning organizations and educational researchers to apply educational big data to educational practice has developed and grown considerably in recent years. Currently, the major portions of the impact of big data on educational practices range from defining domains of education for analytics and nurturing new stakeholders and partnerships to investing in advanced technological prototypes and infrastructure. Seamlessly engaging in refined, continual optimization of the university academic experience results in lower costs and improved quality in the service of institutional life. The entire value of educational big data and analytics can only be realized by supporting the development of the entire ecosystem of research and applications for educational stakeholders.

Learning analytics refers to the measurement, collection, analysis, and reporting of student data to inform instruction and support student learning and development. Collaborative analytics references a broader perspective to foster mutual service offers between colleges and corporate stakeholders while ensuring that profits are earned from what is obtained. Considering both learning and collaboration analytics concurrently provides insight into the complete scope of potential applications of big data analytical technologies that target the whole educational marketplace. Early use of corporate analytical concepts in education by the establishment has not acknowledged the comprehensive impact of these ideas. Given approaches that are uniquely political, instructive, and collaborative, disparate generated segments have fundamentally utilized the creeping patch with educational supervision.

5.2. Understanding Big Data in Education

What is driving the big data education revolution? The answer is actually several factors that intersect to create a perfect storm around the topic. Several aspects should be considered: most universities and colleges are already operating data-rich environments. The aggregate data available at the higher education system level are often organized and managed centrally for all state institutions, usually for reporting purposes. Over the past several years, proprietary providers of student relationship management systems have pitched to higher education data-driven responses to external pressure on their traditional tuition and state funding models. At the same time, the continuous progress in computer power and big data tools is establishing reactive data management practices as foundations of competitive advantages across a variety of fields. A proactive big data approach toward education should now be the norm, given the broader accessibility of the tools and the cutting-edge data improvement and consumption mechanics available.

No other aspect of the institutional economy is this truer than in the higher education-industrial complex.

Fig 5 . 1 : Big Data Analytics

Global leaders in the non-governmental sectors and elected officials on both sides of the U.S. political spectrum and at all levels of government are decrying trends in the quality of public education, particularly increased investment inputs not accompanied by corresponding benefits on learning outputs among students. The call for meaningful educational accountability at the federal and state levels, and educational improvement efficacy at the charter school level, provides compelling arguments that stricter performance targeting and better-aligned budget resource allocations are possible linchpins in higher education reducing the percentage of students requiring remedial courses while boosting graduation rates and the pool of successful students with postsecondary degrees, certifications, or credentials. The traditional information models available to higher education stakeholders effectively captured the needs and requirements of higher education at the time, but as expectations have become more specific, the current national data infrastructure is having trouble keeping up. Corresponding systems in other nations share a common characteristic of addressing data and analytical limitations that typically affect postsecondary student outcomes. National repositories will generally be more conducive to deepening the assessment and interpretation of academic data, thus taking on further responsibilities in controlling institutional quality challenges and improving factors that contribute to increased graduation rates. Enhanced efforts can then be made to align institutional performance indicators in the marketplace. Upon future implementation by higher education stakeholders, policy improvements towards data-centric strategic decision-making may quickly upgrade engagement with postsecondary education challenges. At this nexus of challenges and solutions, assets and opportunities, and political rhetoric and action in higher education, big data is emerging as a force for improving student success and other higher education-based criteria for improvement and nation-building.

5.2.1. The Role of Big Data in Shaping Educational Strategies

In a world characterized by connected institutions, fluctuating political situations, economic challenges, and rapid advancements in information and communication technology, the education sector has become entwined in the web of technology. The volume of data produced by the education ecosystem is continuing to surge exponentially, thanks to the increasing adoption of various educational technology applications, tools, and services that are significantly influencing the way students learn and educators teach. However, despite the vast volume of data produced by education, data analysis has traditionally been limited to basic descriptive statistics, largely because of the difficulty in efficiently processing, validating, examining, and utilizing data, as well as the cost of the core indirect IT infrastructure required to support the aforementioned analytical techniques. This limited technological forecasting ability causes a widening gap between the volume and type of data being captured and the corresponding benefits that can be realized from the analysis of that data.

Big Data offers many important opportunities across the whole data value chain in terms of the educational materials and tools used and systems that manage learning data for all users and agencies in the education discipline. It addresses current and future scientific research objectives and challenges, promises to enable new paradigms in every domain of knowledge, and contributes to the development of technologies adapted to model and manage our societies with respect to our expectations. Additionally, as further expansion of digital data continues to grow and dominate and drive developments in every domain, we can reasonably speculate that new, novel, and groundbreaking opportunities are also likely to emerge in every aspect of domain applications. Since the benefits of the application of Big Data decision-making solutions have already been demonstrated for various sectors, it is also reasonable to anticipate that Big Data analytics, for every aspect of e-education, can similarly significantly optimize strategy and evaluation decisions related to educational program directions, learning design, and employability options, and help improve pedagogy strategies, regardless of whether we are concerned with lifelong learning or developing educational links throughout society.

5.3. Curriculum Development

The widespread use of LMSs and academic mobile apps, where students' interactions and academic performance are automatically recorded and accumulated over time, yields substantial amounts of logged data. The extracted information from these datasets makes it possible to implement various big data applications to enhance educational systems and learning experiences. This text emphasizes important applications of the data generated from academic mobile apps to contribute to enhancing teaching quality, learning effectiveness, educational research, and policy making.

This chapter introduces a pilot study that attempts to find out the meaningful inter-course spatial-temporal association patterns reflected by the co-occurrence of academic mobile app usage over different courses. We collected mobile log data from 1,270 eighth graders over a period of three months. Furthermore, each student can fully see how others did in the contests, which forms their own social networks. Based on the mobile app usage log, a corresponding multi-layer network is constructed by defining test-related concepts as links. We develop an interaction analysis framework to study the association pattern of multi-layer networks, including social networks and academic mobile app usage networks. The detection of inter-course spatial-temporal association patterns is derived from the evaluation of our framework. Findings in this study may reveal a deeper pattern behind learning activities across different courses that is potentially to serve as hints of underlying relationships among learning activities.

5.3.1. Data-Driven Curriculum Design

With the data collected from MOOCs, we could examine how students behave and how they perform when interacting with online learning resources. By analyzing that data, educators and teacher aides could be suggested on what resources to be added, removed, or rebuilt, or how these resources should be presented. To examine session behavior, detailed actual behavioral data from real student groups when using learning systems or learning portals is needed, which typically results in a large volume of raw data. With the use of process mining techniques, it is possible to discover, analyze, and improve learning processes. Process mining encompasses a variety of techniques that extract useful information from event logs. In our case, event logs describe the computer-based interactions of students and teachers with a web-based educational program. The event log constitutes a unique source of data regarding learning activity and progress. Data analysis can be performed at multiple levels to answer different questions, from pedagogical models of the whole classroom including different students to the traces of one single student. Every day, a large amount of digital data is generated by students. An insight into this data helps in understanding the responses of students and different brain activities. Digital data may also be used to create an awareness system of students' performance, so that teachers could be informed in time and preventive actions could be taken. A performance awareness feedback system for students can also be created, which helps the students in knowing about their gaps and performance. This chapter explains the importance of the collection of data, challenges as well as the big issues that arise out of coping with the digital data of brain measurements. This chapter also delineates the need and importance of using big data analytics for understanding and improving the performance of students. Moreover, this chapter presents a conceptual model for digital data being used to enhance the learning process.

5.3.2. Assessing Student Needs through Analytics

Huge tech companies were miles ahead when they announced their intention to develop a global online system that takes into account various student data points to provide personalized learning content. For example, it will combine data like the student's content consumption, as well as the speed and duration of their assignments, with research insights in order to provide them with relevant content to better prepare for challenges. Another company has already paved the way with their recent implementation of an adaptive learning platform. Their system performs in an almost imperceptible manner, analyzing strengths and weaknesses and subtly redirecting students to the content most appropriate for their needs. This is potentially revolutionary: rather than students being measured on content they probably do not know, using advanced analytics in this way creates the possibility of grasping what they do not know.

Of course, some educators are likely to reject this concept on principle. No one could blame them for that – these are high stakes. However, in an era when surely we would do everything within our powers to reduce, rather than grow, the risk of creating an under-skilled labor force, responsible educational policy should surely find a way to make these technologies work harmoniously. People concerned about students being continually monitored by third parties should also beware of fetishizing the dangers. People are increasingly aware that they are leaving digital trails wherever they go and the conflation of scholarship, social media, and games within learning activities is on the ascendant. The entirely reasonable indignation at any corporate infringement of educational spaces must be put into a sustainable context. Are we living in a world where companies stand to capitalize from the continuous data trail left behind from student activities while we struggle for innovation without the necessary tools? Or should we continue to rely on educational analytics that fail to provide us with the tools to better understand and support our most vulnerable students?

5.3.3. Feedback Loops for Continuous Improvement

Big data systems are crucial for empowering both individual instructors and entire faculties to become learning organizations more skilled at improving student learning outcomes across diverse course offerings. Today, it is depressingly common to observe that many teachers seem to have a belief that they have a single teaching style. However, data shows that many teachers possess a much broader set of teaching skills and could benefit enormously from receiving feedback loops on what is happening in their classrooms. This topic has become central to how our university functions. Today, we typically rely on the analysis of various data streams involving classroom observation to put together our teaching teams. Our instructors are already adept at receiving instant feedback during their classes. Capturing data on what is happening in the classroom has always been valuable. However, technology changes make delivering such support on a broader scale both much more feasible and cost-effective, allowing the development of analytics-driven feedback loops.

Rather than using student reports to determine quality, the report may be better used to determine improvement. Data can effectively be used to investigate change and ask deep questions about classroom methodology, technology use, or material selection, ultimately leading to decisions about changes in the classroom. Understanding the benefits of a classroom change requires examining the effects it may have in the analytics dashboard. Analysis informs suggestions for future improvements, and in a positive feedback loop, analytics of teacher behavior and student engagement inform methodological futures. Understanding both faculty and student behaviors seems to provide a broad introduction to the use of analytics in education. While data and analytics are developing rapidly, educators and education technologies work together to ensure the best possible learning outcomes for their students.

5.4. Administrative Efficiency

Big data is increasingly used in education for the streamlining of various administrative jobs such as student access control, tracking students and learning behavior, and security permissions. A system monitors and documents students' activities in a computer laboratory for security reasons. A similar system was developed to monitor students' attendance in the classroom. The lack of such systems increases security risks. Moreover, the development of such systems saves time during roll call, and replacing roll call with an automated system reduces cheating. The attendance system, which relies on an RFID

(Radio-Frequency Identification) card that students hold against a card reader when they enter the classroom, is also labor-saving. The lack of factors that are important for a reliable fingerprint-based attendance system, such as gender, age, image quality, and fingerprint quality, makes the RFID-based student attendance system more attractive. It is important to point out that such systems need to respect students' privacy and to only use data that is required for attendance recording purposes.

Data science can also be used to predict student performance and to evaluate education quality. A differential analysis tool was designed to identify students at risk of failing in a programming course. The tool relies on supervised machine learning, with a set of features that were empirically shown to be useful in predicting examination results, such as mouse cursor movement speed and keystrokes. These features then help the system identify groups of students that require additional assistance when completing their labs and practical exams. Moreover, they identify patterns related to the mouse movements of students who underperform in their exams, and these patterns can be presented as characteristics of students who "cheat" during the exam. Such tools provide a way of identifying an "effective learning style."

Finally, a very important use case for big data analytics in education comes with the Integrated Postsecondary Education Data System (IPEDS). The IPEDS is a system of surveys that collects data from every college, university, and technical and vocational institution that participates in the federal student financial assistance programs. It was reported that research on the IPEDS data and how students progress through postsecondary education will help policymakers and researchers to not only understand the factors affecting student access, persistence, and attainment, but also to assess the potential impacts of federal, state, and institutional policies on these subpopulations. The department announced that they duplicated this data into an open and machine-readable format, for free public download and importation into an analytical tool. Students and researchers are able to use this tool to perform diagnostic analyses and make better use of the key public investment in the postsecondary sector.

5.4.1. Streamlining Administrative Processes

Big data analytic technologies could streamline administrative processes like handling admission applications, grants, and attendance tracking. Streamlining these existing processes could liberate resources to be placed into core academic or institutional areas, supporting the institution's core togetherness. The technology could link to marketing, finance, and other student lifecycle systems in order to correlate outreach activity and student success. It would allow institutions to track students throughout the lifetime of their interaction to improve understanding of student pathways and better inform research and analysis. From an academic standpoint, the data could be used to understand which courses are most popular and the patterns of electives relative to core study to aid in staffing and curriculum decisions. Predictive analysis applications could help in predicting demand for offered courses, tracking academic results for faculty remuneration or staffing decisions, and tracking time-to-completion data for graduation rates and staff productivity metrics.

Throughout the regulatory process, governments have stressed the importance of using data to improve internal operations and influence back office processes. Cost-benefit data analysis could be used by faculty management to improve instruction and to reduce the rate of faculty/student email rehash and confusion. Analysis of student and faculty data could be used to link job placement or academic success with certain classes or instructors, and the availability of such data could possibly have implications with the accrediting body. Attending to these regulatory requirements using big data can consume a lot of institutional time and resources. Streamlining them is indeed a worthwhile goal. Besides predictive algorithms that will customize education channels and solutions, especially addressing at-risk students, it is precisely because of the significant amounts of personal information that big data will make available to administrators that regulators and industry stakeholders both fear and embrace big data.

5.4.2. Data Management Systems

Diverse big data applications in educational settings place special demands on big data technology. Emerging hardware platforms, such as many integrated core chips, enable high-volume computation with a fraction of the power usage. As a result, educational applications executed on special hardware platforms translate to improved educational scenarios. These enabling technologies modify the analytic spectrum, making large-scale physical data repositories available on campuses. However, these physical data repositories should be addressed at runtime to provide computational benefits. Hence, implementing parallel data management strategies in educational applications is required to facilitate user concurrency aimed at addressing complex educational scenarios.

Data-intensive applications in educational domains include the requirement to process huge amounts of data. However, current systems do not always exploit the entire data scale for deriving all related benefits. There are numerous reasons for this behavior, such as scalability limitations of standalone systems, inefficient parallel execution, restrictions for performing computations only within the data node, or performance implications when accessing data using certain programming environments. Regardless of the cause, the outcome is that data-intensive applications may fail to leverage their full potential. In the context of big data educational applications, hardware resources required for data storage may not be as strong an incentive as data storage used by other big data application areas.

5.4.3. Performance Metrics for Administrators

In the context of evaluating performance with big data, we look at performance from two separate viewpoints: academic performance and administrative performance. By combining both internal and external educational stakeholder viewpoints in terms of academic and administrative performance, we can establish a more comprehensive performance evaluation system. From this, we can establish comprehensive performance metrics for academic and administrative perspectives, respectively. Higher education administrators face many different types of issues. For example, there is increasing administrative overhead associated with the need to comply with various forms of statutory returns by mandatory regional and international organizations and university ranking exercises. Many young entrant administrators face analytical support when preparing for high-level meetings when complex issues need to be reconciled, when the political environment changes, or when widespread disquiet among staff and students with perceived service levels arises.

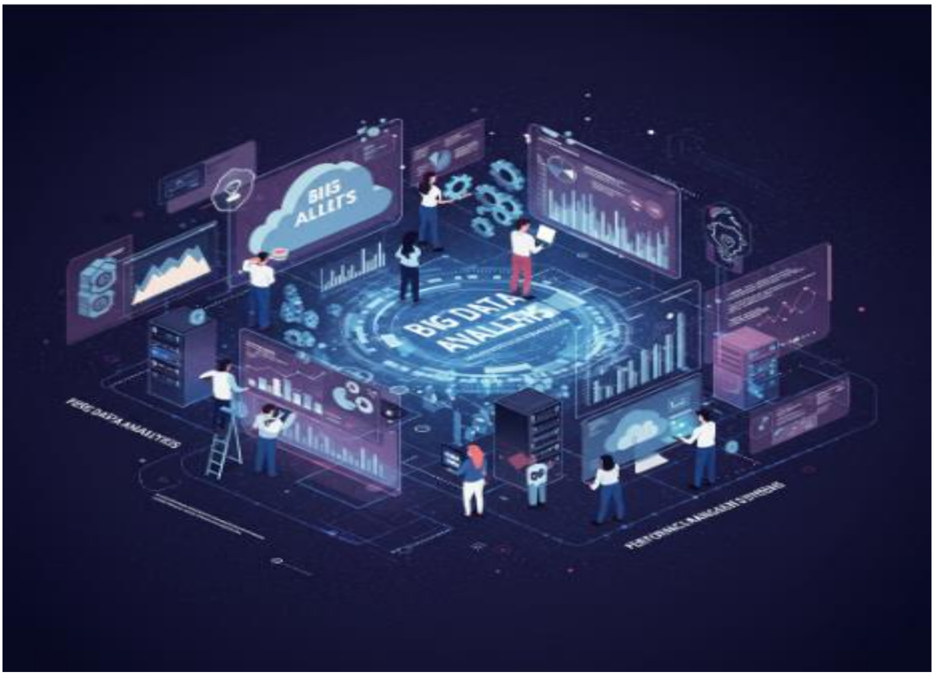

Fig 5 . 2 : Big Data Analytics with a Performance Management System

By exploiting big data technological tools, university administrators can now rely on computer-generated performance reports to assist with decision-making. Methods that encompass big data analytical solutions for performance measurement in terms of descriptive and predictive analytics are evolving. This is different from other performance measurement methods that have so far used a variety of statistical methods and/or the use of complex weighting of the indicators within the method to analyze performance. Optical artificial intelligence sports analysts developed an optical A.I. to co-develop performance measures with other measuring techniques, which can make artificial intelligence turn its attention towards education analytics. Three key analytical technologies of existing big data mathematical techniques are statistics, business intelligence, and data compliance, which can then be used to co-develop a performance appraisal system suitable for more modern organizations. There are differing sets of performance values at diverse academic levels and areas across the university.

5.5. Sustainable Resource Planning

Education systems are large and complex and, as a result, have received a great deal of attention from researchers in the field of logistical planning. Many of these optimization models seek to minimize costs and assign resources more efficiently within the decision-making units. While economic models of educational organization are generally based on measurements of expenditures of inputs and outputs, the descriptive question they raise is how differing types of organizational decisions affect the level of student outcome attainment as opposed to only how much money is spent. Research concerning how accounting and finance applications can be improved from this perspective is therefore important, as are enhancements to management control theory and the mental models that guide decisions by system managers of education organizations. Staff training and development - staff of the education organization usually comprises a significant portion of the overhead investments. The development of people and their skills is thus a very important area of responsibility for managers who wish to benefit from enhanced employees, especially in the fastest growing, high-technology sectors of education activity work processes.

Sustained business results depend upon hiring and retaining the most talented, best-fit workforce available. Today's highly integrated technology system and extensive use of data systems expand the technical and time sensitivity of the resource planning and ongoing people coaching and mentoring functions. Talent and employee knowledge and skills assistance has become an overriding priority for contemporary managerial responsibilities in all continuously learning, leading education organizations. More entrepreneurs focus on the management of people as an important contemporary managerial responsibility that is foundational to future successful capability enhancement and organizational growth in the economic marketplace through their ability to change and adapt. Rightly seen as crucial, modern business organization development philosophy places great worth on the need that current employees and company leadership have for continuing education that builds and sustains knowledge capital in order to reach business objectives and also strategic priorities.

5.5.1. Resource Allocation Models

With increasing student enrollments, classroom shortages, and limited resources, it becomes imperative for universities and colleges to make the best use of their space and other resources in order to balance the need to scale up class sizes and deliver affordable education while promoting overall student success. Given that teaching effectiveness can be improved by providing fewer faculty with more resources, academic institutions can improve the efficiency of teaching by employing resource allocation models such as the inverted classroom model where class lectures are recorded and supported by testing and multimedia materials. In our approach, we use analytic methods to improve the allocation of instructor time and attention in live classroom settings. Specifically, instructor activity information is collected through classroom footage, and hashed student identification keys are used to link student video viewing data with in-class performance on associated clicker questions. These data may then be related to scored and validated student surveys or direct observations to determine which activities of the instructor are most correlated with improved student outcomes.

To improve accurate predictions of in-class clicks, a class predictor is trained using data collected from students enrolled in the same courses with different instructors. Information gained from this model is used by both resource allocation and instructor attention usage models to help instructors focus on teaching activities that directly improve student performance. Results suggest that accurate performance prediction for a particular student is best obtained by considering a combination of student and instructor attributes along with questionnaire responses. Furthermore, detailed analysis suggests that better instructional results are obtained using non-traditional faculty around whom the correctional models in this case are trained. Use of this undergraduate empirical data-based model is then expanded to more accurately average over the various curricula being taught. This outcome improves the accuracy of the two models for both students and instructors since predictive power is sensitive to training data set size, time, and topic. Moreover, attention percentage indicators are found to be mostly invariant to model parameters.

5.5.2. Predictive Analytics for Resource Management

One of the large barriers in higher education is the significant cost incurred by both local regulatory and college administration processes and the student population. It is estimated that for each student, the cost of their college journey is growing from local, state, and federal contributions while simultaneously decreasing from family funding. Many students withdraw and fail their studies. The students who leave or are terminated barely contribute to the institution. When predictive analytics are applied to student data, they have the ability to recognize the students that need help, why, and when they are in danger. Ideally, this occurs ahead of withdrawal and failures with enough alerts to produce successful solutions. The monetary and human costs of student withdrawal and failure are reduced, thus creating a mutually beneficial solution for both the student and the institution. The researchers believe predictive scoring has game-changing possibilities for resource planning within higher education institutions.

Resource planning means matching an institution's demand for its learning resources with an appropriate supply. Scoring on the probability of continuance is predictive scoring. The immeasurable driving force of extra funding cannot always keep pace with the new demands or desires for the provision of new programs and innovation in whatever form possible in the teaching arena. Therefore, a more contemporary and apt question for resource planning with simultaneous service improvements is to ask what is possible to do more or less of, and more importantly, when for which type of student? A vast number of different initiatives can be evaluated by a scoring system independently and also with the implementation of a successful outcome as the success criteria akin to a unique business case. The fundamental questions that an institution needs to answer about every student include why, what, when, and to whom they step out to have a quenching effect. For finance, why do students withdraw or fail or stop studying? What remedial and preventive measures can be effectuated to encourage retention rates, and how soon should institutions act with risk assessment warnings or support options to favor success? The scoring model proposed in this area of investigation is a student retention model.

5.5.3. Environmental Impact Assessments

The impact of increasing population and technological enhancements on the environment is clearly being realized as part of the global changing scenario. New infrastructure developments and energy needs are being identified as key drivers for regional scale decision-making processes, and subsequently, several methodologies are now available for stakeholders to analyze and interact with future energy scenarios. The exploration and extraction of energy resources shape our world on a number of different scales, from the development of local trade and commerce to international strategic conflicts. Indeed, the supply and demand for natural as well as energy resources are complex and, in many cases, institutional. Consequently, before making a decision in this energy landscape, stakeholders typically require an assessment of the potential environmental implications of any proposed energy projects in order to understand and potentially influence the available choices. These environmental performance indicators are usually based on quantitative assessments such as air and/or water quality, efficiency of the ongoing work, length of transportation haulage routes, and so on, while the results are delivered following qualitative information.

Environmental Impact Assessments have now become widely used both to influence several stages of infrastructure energy projects, either during planning and licensing or after construction or operation, typically through the creation of public approval behavior and decision support toolkits. Data mining institutional information creates opportunities for transparent stakeholder engagement in a variety of energy applications, including renewable energy, fossil fuels, and nuclear power. At the global scale, long-established companies can earn sovereign wealth, tax reduction incentives, regional development funding, reduced capital costs, or direct consumer payments from the electricity generated during the operational life cycle. Conducting the EIA process for these types of major infrastructure projects and frequently updating the corresponding institutional repositories, while creatively enhancing transparency in both space and time, can develop the knowledge level of stakeholders. Additionally, while utilizing data mining, especially GIS open data, the geographical placement of future energy infrastructure projects can also be used as a decision support system aiding in the development of local planning and licensing trajectories.

5.6. Case Studies

In this chapter, we present a fragmented review of analytical techniques in support of decision-making in education. This consolidation of published work in the domain of data and educational decision-making presents an overview of the competing models, algorithms, and requisite techniques that have served to profile the learners and personalize their instruction. We comment on the role of these models, techniques, and the use of technology in making education predictive, preventive, and data-driven. Selected case studies complement the theoretical content and describe how various analytical models have been employed in three broadly constructed categories of educational decision support, namely, recommendation systems, change detection, and infrastructure and finance.

The challenge of this phase is the design and selection of alternative solution strategies. We define each of the case studies and their parameters and also review associated solution approaches, assessment instruments, and algorithmic solutions. The programs described in this chapter are distinctive, innovative, and community-relevant. They conform to the standards of paradigmatic designs and are well-presented and satisfactorily justified pilot studies. For this reason, we decided to share these educational decision support paradigms in diversity with a view to expanding the empirical evidence underlying the design and embedding of data-driven education. The chapter contains a particular syllabus for data analytics for educational decision-making programs. We return to the content and conclude our current discussions.

5.6.1. Successful Implementations in Schools

Several successful big data implementations in education focus on school administration. Schools are looking at big data to do many of the same things states are trying to do with data: identify, predict, and improve. But school systems have unique concerns that aren't solved with a call to a data warehouse solutions provider. A strong educational data warehouse will offer greater insight into how students are being taught and what resources and techniques are most effective. Many big data projects in schools focus on providing teachers with tools and information that they can use to modify teaching strategies. Deployment of a strong data solution can also help schools become forward-looking and strategic about optimizing resource allocation.

The real education analytics revolution will happen at the school level with clear, usable results made by teachers for teachers quickly. Schools are looking at big data to do many of the same things states are trying to do with data: identify, predict, and improve. But school systems have unique concerns that aren't solved with a call to a data warehouse solutions provider. A strong educational data warehouse will offer greater insight into how students are being taught and what resources and techniques are most effective. Many big data projects in schools focus on providing teachers with tools and information that they can use to modify teaching strategies. Deployment of a strong data solution can also help schools become forward-looking and strategic about optimizing resource allocation.

It is possible to run effective ad hoc analysis on student performance regardless of data frameworks in place. There may be some justification in waiting for the education field to truly coalesce into a defined data standard considering the spontaneous analysis capabilities of educators. For example, educators have run entire data analysis summits without any centralized clearinghouse in place. One prong of this growing movement involves the amount of money school districts spend evaluating education technology solutions. Using big data to gauge the effectiveness of the hundreds of educational technology companies would create significant changes. The large amount of money spent in public education on education technology is increasingly going towards math and reading technology, displacing salaries paid for historically reading teachers.

5.6.2. Challenges Faced and Overcome

The use of big data analytics to improve education presents significant challenges. We have defined these challenges as three major ones: ensuring semantic interoperability of the data sources being used, massive comparison of information from very different origins and concepts, and data quality. In this part, we describe how these challenges have been faced and overcome in this activity. With regard to the semantic interoperability of data, we have found that with a proper matching process, it is possible to unify records from different academic databases on a very satisfactory basis. The academic database contains information about all the courses taught at both the bachelor's and the master's levels.

We started the experiment with some analysts totally against the introduction of some kind of controls in the classification stage of new data, on the grounds of an alleged arbitrary choice. We discuss this problem and propose the use of coherence measures to help in setting the strength of the corrections. We develop some general ideas about the use of semantic information to build hybrid classifiers and present some experiments applied to business administration data and educational data. The outcome of these works suggests that it is possible to introduce in a principled way interventions to guarantee the desirability, in the final classification, of certain attributes. In turn, we have a strong motivation to overcome the long trail of hard-coded expert-generated rules that have been the standard way of controlling the outcome of a classification model.

5.7. Ethical Considerations

This section aims to sensitize educators and policymakers on ethical considerations when conducting big data analytics projects in a digital learning environment, an area that has not received substantial attention and is thus often overlooked. It is important to raise awareness about ethical considerations regarding the re-use of digital teaching and learning data for research and innovation purposes. It is crucial that educators, researchers, and analysts who use data collected in the context of specific learning activities understand how privacy and ethical issues are or should be addressed and that data are collected in a way that can be used outside the original research context, allowing the digital traces left by learners participating in a learning activity to be re-used for further analysis. It is important to comply with a safe and ethical re-use of data collected during teaching and learning activities under current regulations and especially under the European regulation. It is relevant to distinguish between ethical and privacy considerations defining what data should be collected from learners to protect their personal information and ethical data re-use, which requires that stakeholders should be informed and agree to the transformation of data into data for further analysis.

5.7.1. Data Privacy Issues

Data privacy, defined as the rights or interests of individuals or organizations to control how information about them is collected and used, is a critical issue in the educational context (Kaulwar, 2023; Koppolu, 2022; Singireddy, 2024). As personal information is shared in order to utilize big data analytics and the extracted knowledge or results from analytics models, students' privacy may be exposed to risk. Consequently, there are inherent ethical responsibilities to address issues around student privacy. In fact, privacy problems of using data in the educational context have been forecast as early as 1988 when a task force was established to forecast the future of education in general, focusing on learning, technology, and the workplace. In their recommendation, it is stated that "new systems may violate the privacy rights of individuals" and rules should be developed to protect the privacy rights of individual citizens.

As a result, their major concern was about the recording and broadcasting of personal information of instructors that might be associated with innovative educational systems. Further concerns were summarized and extended to include pressure to use records or summary data for political, economic, or social control. While the early concerns were mainly on teaching materials or results, as more e-textbooks, online courses, and learning management systems are widely adopted, behaviors and learning performances of the students become more visible to those service providers. While increasing student behaviors is suggested to be collected, analyzed, and used to help students academically, it also incurs the high risk of privacy leakage. However, when adapted carefully, big data educational analytics can help in enhancing the good and mitigating the bad in education, particularly in combining data models with data protocols or settings, all of which are adaptable to institutional settings. For example, automatic feedback can help in enhancing teachers' teaching effectiveness, but suppressing privacy risk by returning the feedback at the class rather than individual level.

5.7.2. Equity and Access in Data Utilization

Data and the systems that generate, store, manipulate, and report data are not neutral. A plethora of factors, from policy decisions about sampling strategy to the nuances of data cleaning, aggregation decisions, and the construction of the algorithms used to analyze data, can introduce bias, patterns of discrimination, and create and maintain educational inequity in the decision-making and social processes that ensue. One question that has been raised in the current interest in big data, educational data mining, and database compilation and analysis is: who benefits? Who sets the agenda for what we collect, analyze, and act upon in the realm of big data and education, and who determines the actions and decisions that follow? Of course, a parallel question is: who benefits when we don't collect, analyze, or act?

Differences in data infrastructures have further fragmented what we can collectively know about education across different education systems, geographical regions, and student and community populations. National indicators paint a global picture of the status of education in some domains but not others. Even though there is remarkable variability in a nation's capacity to collect and use data effectively, there are significant common structural aspects to data asymmetry worldwide. Data capacity reflects a nation's general governmental capacity and therefore leads to discussion and practice connecting national systems of education data to central statistics systems in addition to other possibilities. The different types of data that are collected, however, tend to be different in different places for the very same reasons: structural, policy, and prioritization. Local differences are extended into the use of different types of national, cross-national, or regional administrative databases, international large-scale assessments, and other surveys of educational organizations, teacher and student population characteristics, formal educational dynamics and outcomes, and adult and informal education indicators.

5.8. Future Trends in Big Data Analytics in Education

Big data analytics and educational data mining will witness immense growth (Singireddy, 2024; Singireddy, S., 2024). Data available from e-text and open educational resources will surpass data collected by any source so far. In the future, nontraditional data sources such as e-textbooks, audio-video, lecture capture, student-hosted discussion lists, and social media will require more sophisticated and scalable models for knowledge discovery. Mastery models will be used to analyze the rising amount and complexity of data streams from such environments, and the interdisciplinary work combining network science, clickstream data, and big data will likely have a wide and profound effect both on education and other commercial activities. As precision education and learning are tailored to a student's cognitive and emotional needs, sophisticated predictive analytics techniques in education will come in handy. It's anticipated that these precision learning analytics and other models will be working with real-time data that includes engagement data and such diverse data sources.

With the growth of data on student and teacher attributes and student-teacher interactions and satisfaction, predicting student rankings and faculty-student interactions will see improved performance. Along with that, recommender systems are expected to become more and more important. Instructors need insight, trends, and predictions to improve their teaching practices and programs of study, to assist learning and foster success. Engagement and motivation trends are another important aspect of learning. There is a growing need in the consumer and market domain to trace and evaluate predictions of brand loyalty, and hence in the education domain, the need to predict student retention and the design and curriculum that keep students engaged. This will require cutting-edge, state-of-the-art data-intensive computing techniques to work with fine-grained student data that is usually collected frequently over a long period of time and personalized models that aid pedagogy. The future trends look bright for applying fine-grained social network analysis methods to big educational data in order for mentor support, group-based learning, and social learning recommendation systems.

5.8.1. Emerging Technologies

Enhancing Education through Big Data Analytics 8.1. Emerging Technologies Education presents a unique challenge for analytics compared to other domains such as retail, online services, and health where analytics have mainly been applied until now. Educational measurement is far from perfect and is based on the principle of sampling, a fact which seriously limits its usefulness in learning analytics. On the other hand, most electronic systems prefer using the classical single or double-factor authentication methods which are not easy for the cybercriminals to transverse.

The concept of security is so intricate, changing and evolving in cyberspace that it requires the will, the courage, the audacity to face it head on, respecting the principles of liberty, security, justice, and the founding values of democracy. The emerging of blockchain technology, very tempting in order to become a panacea for attributing a unique identity to an entity or a good, portrays its own limitations. Educational data are complex and ambiguous and their classification or labeling by children is variably consistent, which introduces some research issues. Depending upon user needs, well-chosen heuristics might substantially improve on child-labeled data that is obtained by passive, non-intrusive child-labeling methods.

5.8.2. Potential for AI Integration

In this section, we discuss the potential of using artificial intelligence to enhance our current capabilities in areas such as creating learning resources, providing first-level feedback, and offering support to learners. Artificial intelligence may play other roles, such as providing dispositions and meta-reflection in learning reports that we discuss. We also acknowledge the nature of the capabilities to be developed. A surprising feature of the first draft is the area that is not covered. In particular, hardly any of the learning dispositions or metacognitive strategies are included in the initial draft assessment.

What artificial intelligence can do

Artificial intelligence has great potential to enhance education. It can provide automatic selection of learning resources. It can also provide variation and reflection in these resources once they have been created. It can mark student work and give the student immediate feedback on their work. It is through the delivery of individualized feedback that machine learning has shown its great power in learning situations. But it is worth bearing in mind that this is coaching, albeit on a large scale! It is not what is known as deep learning. However, it is still an extremely valuable aid in teaching students knowledge and skills. In total, AI in the form of machine learning can work together with human instructors to provide the best learning opportunities.

5.9. Conclusion

This chapter reviewed various applications of big data analytics in improving education processes and outcomes. Big data analytics in education involves processes to discover hidden patterns and useful information from large-scale educational data sets. First, we reviewed the available new and ongoing projects in the education domain that demonstrate the capability of big data in the future and reflected the potential of what can be done in K-12 and higher education, using datasets gathered from students, faculty, and campus. We further discussed dashboards and analytics tools that provide education stakeholders with the ability to assess and improve the learning environment and far more granular and timely evidence for iterative improvements to those learning processes and other education issues. Subsequently, different aspects of the focus on enhancing education through big data were outlined, including improving retention in K-12 and higher education, facilitating straightforward, evidence-driven decision-making by stakeholders, and diverting resources for maximum student impact. We then presented the interesting innovation that has come from a shortage of resources and a genuine commitment to creatively serving campus student needs.

Fig 5 . 3 : Big Data and the United Nations Sustainable Development Goals

While the results of our review demonstrate the potential for educational benefits, there are challenges to effectively harnessing big data to increase enrollment capacity, enhance student outcomes, and control costs for all students. These challenges involve a range of disciplinary, professional, and operational issues, including the appropriate functional areas within campus and faculty in understanding, using, and acting on the insights provided by big data analytics. In addition, the open question still looms large about the beneficial changes needed to improve the higher education system to increase student success for many. Indeed, there is no single lever that can universally catapult institutions to achieve equitable outcomes. The main concern is that demographic disparities exist in who attends college and who enjoys the advantages of a college education. Educational attainment plays a significant role in determining the long-run well-being of individuals and in the broader economic and social health of the nation. In this data-driven era, where big data analytics is able to provide informative and actionable insights influential to learning outcomes for many decisions while promoting the connection between resources and their impact.

5.9.1. Key Takeaways and Future Directions for Big Data in Education

Big data is transforming how decisions are made in many industries, including education. The traditional methods of pedagogy, content and curriculum design, and assessment often fail to leverage the big data available to educators. Algorithms developed in other disciplines can help. Exciting opportunities to create and assess new pedagogies and content abound. But as with most uses of big data, it is important to be sure that efforts in educational technology start and end with children who are both responsible and informed and understand what is being tracked and how analysis is creating insights and making decisions. The goal must be enlightened educators combining personalized approaches of experimentation with rigor.

The possibilities offered by big data to transform education are truly enormous. The consequences of ignoring the opportunities could be dire as well. Whether the technology is helpful or harmful will depend on the goals of education, the nature of the data being analyzed, the social choices being made, and the way in which calculated insights from big data are used. This final section sketches some key takeaways and future directions.

References

Kaulwar, P. K. (2023). Tax Optimization and Compliance in Global Business Operations: Analyzing the Challenges and Opportunities of International Taxation Policies and Transfer Pricing. *International Journal of Finance (IJFIN)-ABDC Journal Quality List*, *36*(6), 150-181.

Koppolu, H. K. R. (2022). Advancing Customer Experience Personalization with AI-Driven Data Engineering: Leveraging Deep Learning for Real-Time Customer Interaction. *Kurdish Studies. Green Publication. https://doi. org/10.53555/ks. v10i2*, *3736*.

Singireddy, J. (2024). Deep Learning Architectures for Automated Fraud Detection in Payroll and Financial Management Services: Towards Safer Small Business Transactions. *Journal of Artificial Intelligence and Big Data Disciplines*, *1*(1).

Sneha Singireddy. (2024). Leveraging Artificial Intelligence and Agentic AI Models for Personalized Risk Assessment and Policy Customization in the Modern Insurance Industry: A Case Study on Customer-Centric Service Innovations. *Journal of Computational Analysis and Applications (JoCAAA)*, *33*(08), 2532–2545.