Title: Leveraging Data Analytics to Enhance Student Retention in Higher Education Institutions

Student Name: Michael Ayomide, FADELE

Student ID NO: 239032281

Date: 11/03/2024

Contents

1. Introduction

Increasing student retention in higher education institutions affects students' academic progress and the institutions' financial and reputational position (Thomas, 2019). Data analytics may help solve these problems by strategically analysing data to inform decision-making. This literature review examines the most recent studies and methodologies for employing data analytics for student retention, concentrating on the most important components of data science product design and development in higher education.

Student retention and dropout rates have long been a problem, and Tinto's model for student retention as illustrated in fig 1 below provides a basis for understanding them (Tinto, 2015). Big data and sophisticated analytics have changed the paradigm, providing new methodologies for detecting at-risk children and designing intervention measures. Data analytics lets institutions analyse massive volumes of data from student information systems, learning management systems, and social media to understand student behaviour, engagement, and academic success (Picciano, 2022).

Using data analytics in higher education is difficult. Data privacy, ethics, and data governance are crucial (Slade and Prinsloo, 2023). Furthermore, data availability and quality determine data analytics success. Institutions must often spend a lot of time cleaning, integrating, and standardizing data from many sources to develop a data repository for analysis (Daniel, 2015). Despite these obstacles, data analytics may improve student retention. Institutions may help at-risk students and improve outcomes by recognizing patterns and determinants of student success and risk (Siemens and Long, 2021). According to Arnold and Pistilli (2022), predictive analytics can predict student performance and indicate the need for assistance before children fall permanently behind.

This literature review critically examines higher education data analytics for student retention. Explore accessible data and data repositories, comprehend corporate decision-making processes driven by data analytics, evaluate software engineering methodologies, and evaluate machine learning models. The review seeks to give a complete overview of existing practices and upcoming developments in the sector, emphasizing the importance of data analytics in driving higher education student retention tactics.

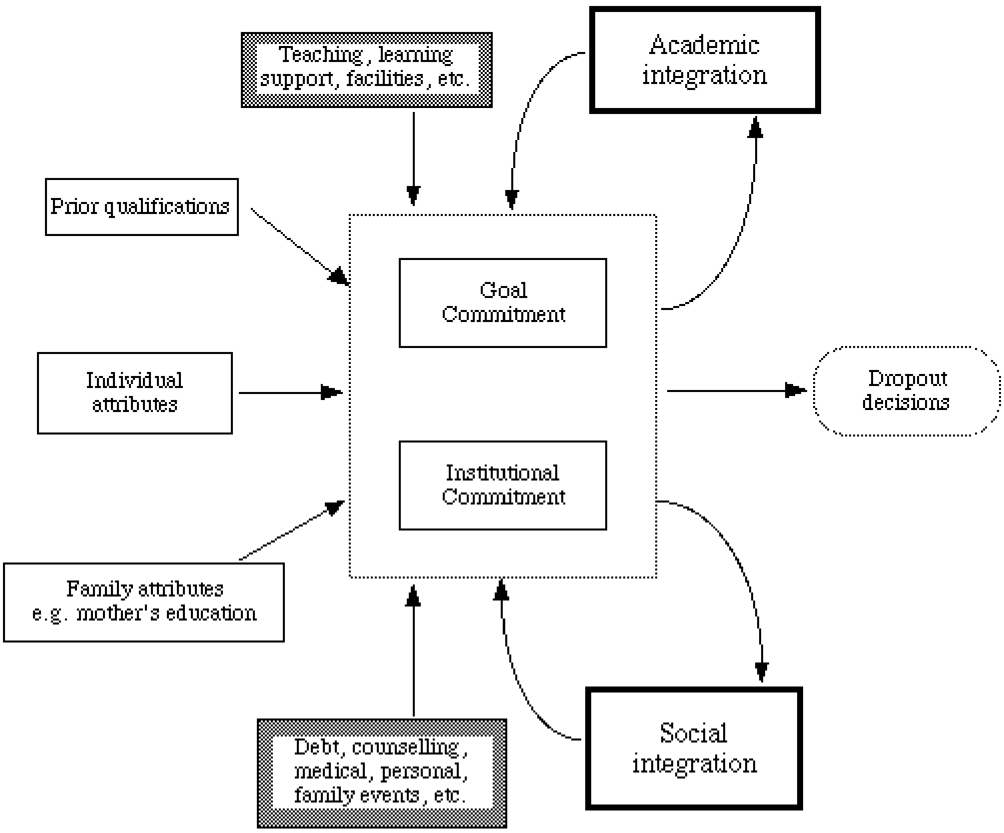

Fig. 1. *Tinto's Model for Student Retention*

2. Review of Related Available Data

Higher education institutions use data analytics to improve student retention. This section examines the many student retention characteristics that institutions gather, including academic performance, engagement measures, and demographic data, as well as the key data repositories and data sets that are necessary for measuring student success.

2.1 Academic Performance

According to York et al. (2019), academic performance data, including grades, course completions, GPA, and degree progress, are crucial markers of student retention. These measures help institutions identify academically failing students and provide tutoring or supplementary instruction. Astin (2017) stresses the predictive usefulness of academic performance in predicting student dropout risk, arguing for early and individualized academic coaching based on these data points.

2.2 Metrics for Engagement

Engagement indicators show students' outside-of-classroom connection with the school. These variables, according to Whitt et al. (2023), include extracurricular activity participation, university resource use, teacher and peer contacts, and online learning engagement. Engaging students leads to higher student satisfaction and retention; hence, institutions should encourage active student involvement (Trowler, 2020).

2.3 Demographic Data

Demographic data is essential for understanding students' different origins and experiences in higher education. This data, which includes age, ethnicity, socioeconomic position, and first-generation college status, helps institutions uncover discrepancies in student outcomes. Pascarella and Terenzini (2020) emphasize the relevance of demographic data in personalizing support services to varied student groups, creating a more inclusive and fairer educational environment.

2.4 Datasets and Data Repositories

The Integrated Postsecondary Education Data System (IPEDS) is essential for higher education data, covering institutional characteristics, student enrolment, financial assistance, and graduation rates. IPEDS, managed by the National Centre for Education Statistics (NCES), is crucial for benchmarking and policy analysis (Schuh, 2022).

The National Student Clearinghouse (NSC) provides precise data on student enrolment, transfer rates, and degree completions among institutions. Shapiro et al. (2017) said the NSC's comprehensive database facilitates longitudinal studies on student trajectories, helping researchers understand retention and completion variables.

National databases, in addition to internal institutional data systems like Learning Management Systems (LMS), closely monitor student engagement and learning outcomes (Vogt, 2016). According to Beer et al. (2022), LMS data can disclose student participation with course materials and online chats, indicating disengagement or academic problems. The landscape of accessible data for studying student performance is further enriched by specialized longitudinal data sets like the Beginning Postsecondary Students Longitudinal Study (BPS) and the Education Longitudinal Study (ELS) of 2022. Ingels et al. (2017) and Chen (2022) follow students across time, giving significant insights into varied student cohorts' educational trajectories and results.

3. Organisational Decision-Making Policy

The changing higher education landscape has influenced institutional policymaking through data analytics. The ability to gather, analyse, and understand massive volumes of data gives higher education managers a strong tool for making decisions that affect student retention and success. Picciano (2022) states that strategic data analytics helps institutions find hidden patterns, trends, and correlations to create focused and successful strategies.

3.1 Informing Policymaking Using Data Analytics

Several significant data analytics channels inform higher education policymaking. It identifies students at risk of leaving school or lagging academically early. Institutions can use predictive analytics to evaluate historical data to identify student attrition risk factors and conduct early intervention methods. According to Arnold and Pistilli (2022), this proactive strategy allows institutions to help at-risk students before they face academic issues.

Additionally, data analytics are crucial for resource allocation and program assessment. Using engagement measures, academic performance data, and other indications, institutions may evaluate which programs and services help students succeed. A data-driven approach to resource allocation ensures that funding goes to projects that improve student outcomes. For institutional policies and actions to be effective, Daniel (2015) recommends smart resource allocation.

Data analytics helps tailor education to student requirements. Institutions can discover differences in student outcomes and establish tailored programs to reduce these gaps by examining demographic data with academic and engagement measures. Bichsel (2022) noted that using data analytics for policymaking helps institutions develop more inclusive and equitable educational environments that benefit all students.

3.2 Decision-Making Frameworks Using Data

Several decision-making models help institutions use data for student retention and achievement. Data-Driven Decision Making (3DM) emphasizes systematic data gathering and analysis across all institutional operations. Mandinach and Gummer (2016) say the 3DM framework promotes evidence-based decision-making, where policies and actions are evaluated and improved based on data. The Integrated Planning and Advising System (IPAS) leverages data analytics to improve academic advising and student assistance. IPAS combines data from several sources to give advisers a complete picture of students' academic achievement, engagement, and risk factors. Personalized coaching and assistance from advisors improve student retention and success rates with this comprehensive approach. IPAS requires higher education institutions to employ data analytics to influence policymaking and intervention initiatives, according to Beer et al. (2022).

Additionally, the Learning Analytics Framework (LAF) provides a data analytics methodology to improve teaching and learning. Siemens (2023) defines LAF as managing student learning behaviours, course interactions, and academic performance data. This data may help create tailored learning routes, instructional tactics, and curricular innovations that boost student engagement and accomplishment.

The incorporation of data analytics into institutional decision-making processes advances higher education institutions' strategic management. Education institutions may improve student retention and achievement by using data to influence policy, resource, and intervention measures. The 3DM, IPAS, and LAF decision-making frameworks incorporate data analytics into institutional operations to provide evidence-based policies that enhance student assistance and academic performance.

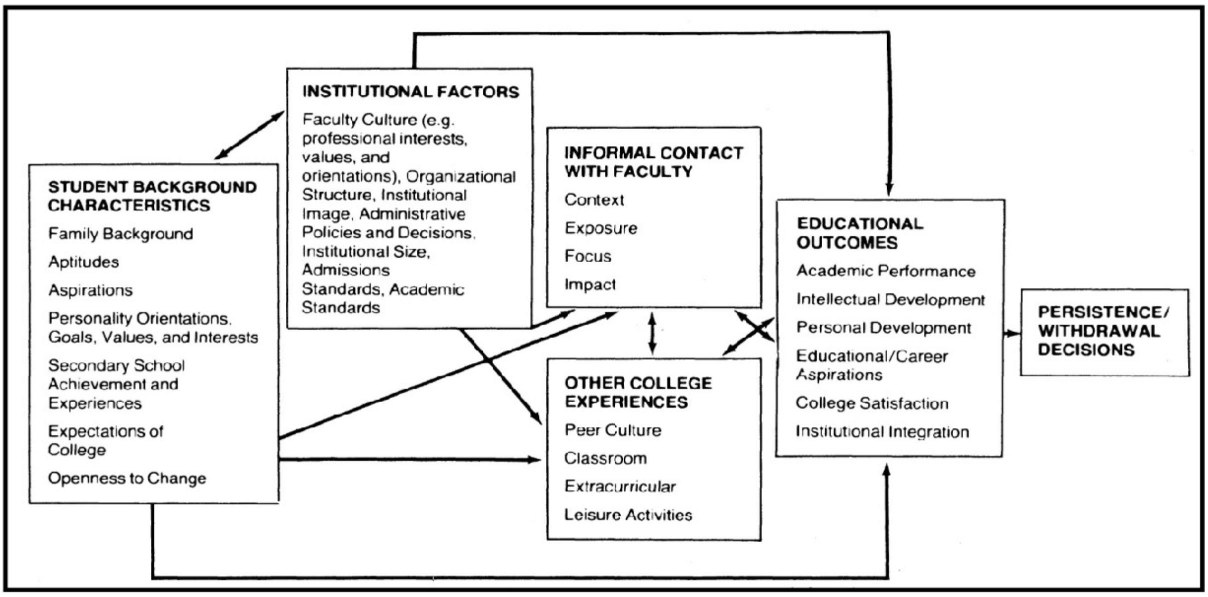

Fig. 2: *Framework showcasing the diverse student and institutional factors influencing retention and complicate policymaking*.

4. Software Engineering Methodologies for Data-Driven Solutions

Software engineering and data science are needed to provide data-driven educational solutions. This method guarantees that analytics tools are technically sound and connected with educational aims.

4.1 Evaluating Software Engineering Methodologies for Educational Data Science Projects

Software engineering approaches include planning, design, implementation, testing, and maintenance. Agile and DevOps are beneficial for education data science projects owing to their flexibility, iterative nature, and focus on collaboration.

The agile approach, with its iterative development cycles and emphasis on adaptation, suits educational data science projects' dynamic needs. According to Talukder (2023), Agile allows project teams to quickly adapt to data analysis-driven insights and needs, keeping solutions current and effective. Agile allows stakeholders, such as educators and administrators, to provide input on project goals and deliverables.

In educational data science projects, DevOps, which combines development and operations teams to boost cooperation and productivity, is also gaining popularity. Demchenko (2020) states that DevOps improves data-driven application deployment and administration, making them scalable, dependable, and secure. In educational contexts, data privacy and security are crucial. In addition, Lal and Kumar (2021) recommend Test-Driven Development (TDD) for data analytics algorithm correctness and dependability. Tests before coding help developers confirm their solutions satisfy requirements and work properly under varied scenarios.

4.2 Assessing Project Management and Development Methods for Analytics Solution Design

Project management and development techniques are essential for data-driven education solutions. For data science projects' particular problems, Martinez et al. (2021) recommend data-centric project management. This requires prioritising data quality, access to relevant data sources, and strong data governance throughout the project lifetime.

Li et al. (2022) say cross-functional teams are essential for software engineering-data science collaboration. These teams use software development, data analysis, machine learning, and educational philosophy to build comprehensive solutions. Collaboration tools and platforms that enable Agile and DevOps let team members communicate and collaborate.

Continuous integration and continuous deployment (CI/CD) pipelines, as advocated by Nogueira and Zenha-Rela (2021), improve data-driven solution development and deployment efficiency. CI/CD approaches automate code testing and deployment, speeding up updates and aligning software with changing educational requirements and data insights. Small projects can grow in complexity and breadth, making scalability important. Houmani (2021) recommends a microservices architecture for data-driven applications, where components are built and delivered separately. Scalability and feature updates without system disruption are simpler with this design.

In addition to technological techniques, ethical and data privacy issues must be addressed early in the project. Tzimas and Demetriadis (2021) state that ethical review processes and privacy impact assessments throughout the project management lifecycle guarantee that data-driven education solutions respect student and instructor privacy and rights.

5. Data Science Methodologies and Machine Learning Models for Student Retention

McNamara et al. (2022) claim that data science and machine learning (ML) models have changed how educational systems deal with student retention. These tools enable data analysis, retention prediction, and personalised interventions for at-risk students. This literature review analyses data science methods for student retention and assesses ML models for dropout prediction, at-risk student identification, and intervention personalisation.

5.1 Analysis of Data Science Methodologies Applied to Student Retention Problems

1. **Predictive Analytics**: This data science strategy is essential for student retention (Syed, 2023). Predictive models can anticipate dropouts by analysing past data and finding patterns. Marbouti et al. (2016) used predictive analytics to uncover online course performance variables, emphasising early engagement metrics.

2. **Descriptive Analytics**: According to Park and Jo (2015), descriptive analytics analyses data to understand prior behaviour. This approach shows institutions how retention tactics are working by identifying dropout rate trends. Varlotta (2016) found that descriptive analytics can help explain how academic advice affects students' perseverance.

3. **Prescriptive Analytics**: Prescriptive analytics suggests activities using predicted data (Poornima and Pushpalatha, 2020). Prescriptive approaches recommend individualised instruction or counselling for dropout-risk students. This method optimises resource allocation to the neediest, improving retention.

5.2 Evaluation of Machine Learning Models for Predicting Student Dropouts

1. **Decision Trees and Random Forests**: These models offer transparent and interpretable models for predicting student dropouts. These models analyse various factors, such as academic performance and engagement levels, to identify students who are at risk. Baradwaj and Pal (2022) demonstrated the effectiveness of decision trees in identifying potential dropouts, highlighting their utility in educational settings.

2. **Neural Networks**: Neural networks, especially deep learning models, are famous for handling complicated, nonlinear interactions. These algorithms can analyse academic records and social integration measures to forecast student retention with subtlety. However, their "black box" aspect might hinder interpretation (Kovačić, 2020).

3. **Support Vector Machines (SVM)**: SVMs are another powerful tool for classification tasks, such as predicting student dropouts (Del Bonifro et al., 2020). High-dimensional SVM models are ideal for analysing datasets with many variables. Osmanbegovic and Suljic (2022) found that their hyperplane-based class separation may detect at-risk students.

5.3 Personalising Interventions with Machine Learning

Machine learning (ML) technologies in education represent an evolutionary leap beyond student academic outcome prediction (Rane et al., 2023). It signals a new age of personalised education where interventions are proactive and tailored to each student's needs and choices.

This revolutionary strategy is transforming education, bringing promise for student retention. By using ML's predictive capacity and flexibility, educators and institutions may spot potential issues and act before students fall behind or disengage. Masika and Jones (2016) claim that this proactive approach to education creates a more supportive and engaging learning environment that motivates students. Student retention rates rise when they feel understood, respected, and supported in their education. According to Kucirkova and Leaton (2023), ML is democratising personalised learning, making it available to more students in different educational contexts. ML technologies are helping all students, regardless of background or learning style, excel academically by removing obstacles to individualised support (Abendan et al., 2023). The educational system becomes more egalitarian and inclusive, and student retention improves.

Integrating certain ML approaches shows this personalised education shift:

1. **Segmentation Through Clustering Algorithms**: Using algorithms like K-means to cluster students by similar qualities helps provide customised assistance measures. Advanced clustering approaches can personalise educational routes, improving retention and academic achievement (Xu et al., 2017).

2. **Customisation with the Recommender System**: These systems tailor recommendations to each student's tastes and requirements using predictive analytics. Recommended systems propose educational materials and activities based on past data on students' interactions, preferences, and performance, according to Drachsler et al. (2015).
3. **Adaptation using adaptive learning systems**: Using student performance assessments, these systems adjust learning material and complexity. Pardo (2019)

adds that these methods are vital for student engagement and retention by maintaining optimal engagement levels.

4. **Sentiment Analysis Insight**: According to Floyd et al. (2021), natural language processing (NLP) analysis of textual feedback reveals students' emotional well-being and contentment. Sentiment analysis can identify students' emotional and academic assistance needs, enabling early and successful interventions, according to Arnold and Pistilli (2022).

Incorporating these ML methods helps education institutions predict at-risk students as well as implement precise, individualised interventions.

6. Implementation and Impact of Data Analytics in Higher Education

Data analytics has transformed student retention methods in higher education. Big data and analytics let institutions quickly identify at-risk students, customise interventions, and optimise resource allocation. These data-centric interventions are examined in case studies of data analytics implementation.

6.1 Illustrative Examples of Data Analytics Deployment

- **Georgia State University (GSU)**: GSU pioneered data analytics to improve student retention and graduation (Salmi and Sursock, 2018). For early dropout risk detection, GSU's "Student Success" programme carefully analysed over 800,000 student data sets using predictive analytics. Renick (2020) found that an early-warning system enabled prompt advising interventions, reducing degree completion time and increasing graduation rates, particularly for historically underprivileged groups.

- **Purdue University**: According to Boser (2021), Purdue University's Signals system, which uses data analytics to dynamically monitor student involvement and academic achievement, has improved. The system proactively informs underachieving students by combining demographic data, coursework grades, and LMS data. Arnold and Pistilli (2012) note that this method has improved course completion and student accomplishment.

6.2 Impact of Data-Centric Interventions

The efficacy of data-driven higher education interventions has received ample documentation. Student engagement, academic performance, and graduation rates increase when institutions use data analytics for student retention. Data analytics' targeted nature allows educators to give personalised help, addressing students' unique issues and enhancing their chances of success, according to Chen and Upah (2020).

6.3 Assessing Scalability and Sustainability

- **Scalability**: Scalability is a key benefit of data-centric interventions. As Sclater et al. (2016) explain, institutions may use cloud-based analytics systems and scalable databases to analyse data from many students across programmes. Scalability guarantees that data-derived insights may benefit many students.
- **Sustainability**: Continuous technology and staff training are needed to sustain data-driven interventions. According to Daniel (2015), this includes frequent software upgrades, staff data analysis training, and data governance regulations to protect privacy and ethical data use. For long-term success, data analytics must be integrated into the institutional culture, favouring evidence-based over intuitive decision-making.

6.4 Navigating Challenges

Higher education data analytics integration confronts several challenges, despite its promise. Challenges including data protection, technological support, and the risks of overusing quantitative analysis demand careful consideration. Siemens (2023) adds that data-driven strategies succeed when institutions can turn findings into responsive, practical plans that meet students' changing requirements.

7. Conclusion

This literature review outlines that data analytics in higher education combines technology and pedagogy to solve student retention problems. Case studies, approaches, and frameworks show that strategic data analytics may improve student performance and institutional effectiveness.

7.1 Summary of Key Findings

1. **Strategic Implementation**: Georgia State University and Purdue University use data analytics to boost student retention, graduation rates, and engagement.
2. **Comprehensive Methodologies**: The review emphasises the need to use predictive, descriptive, and prescriptive analytics and machine learning models like decision trees, neural networks, and SVMs to predict and shape student outcomes.
3. **Personalised Interventions**: ML technologies enable a new age of personalised education, where interventions are matched to students' needs and preferences, improving educational performance.
4. **Challenges and Considerations**: Data analytics has significant potential, but the review notes data privacy issues, the necessity for constant technological assistance,

and the risk of over-reliance on quantitative indicators. It emphasises ethics and a data-informed institutional culture.

7.2 Reflections on the Potential of Data Analytics

Data analytics might revolutionise higher education student retention. By studying student behaviour, engagement, and performance, institutions may transition from reactive to proactive tactics that address present and future concerns. This trend towards data-informed education can democratise education by adapting the learning experience to the various requirements of students.

7.3 Recommendations for Future Research Directions and Policy Implementation

1. **Future Research**: A future study should examine the long-term effects of data analytics on student retention, focusing on personalised interventions across demographic groups and educational experiences.
2. **Develop Policy**: Data governance, privacy, and ethical analytics regulations should protect students' rights while maximising data-driven tactics.
3. **Training Investment**: Educational and administrative workers must receive ongoing professional development. Data analytics training should emphasise both technical and pedagogical applications to promote data-driven insights.
4. **Collaboration and Sharing**: Sharing best practices, insights, and problems amongst institutions may create a community of learning that promotes higher education data analytics adoption.
5. **Assessment/Adaptation**: Data analytics programmes need solid systems for continual review. To be relevant and effective in a changing educational world, feedback loops should influence strategy adaptation and revision.

In conclusion, the integration of data analytics into higher education represents a promising frontier for enhancing student retention. By navigating its challenges thoughtfully and leveraging its full potential, institutions can unlock new pathways to student success, making higher education more responsive, engaging, and inclusive.

REFERENCES

Abendan, C.F., Kilag, O.K., Taping, M.L., Poloyapoy, G., Echavez, R. and Suba-an, H.M., (2023). Driving Excellence in Management Education through IT Innovation. *Excellencia: International Multi-disciplinary Journal of Education (2994-9521)*, *1*(4), pp.62-75.

Arnold, K.E. and Pistilli, M.D., (2022), April. Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270).

Astin, A.W., (2017). *What matters in college?*. JB.

Baradwaj, B.K. and Pal, S., (2022). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*.

Beer, C., Jones, D. and Clark, D., (2022), January. Analytics and complexity: Learning and leading for the future. In *Proceedings of the 29th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2012)*. University of Southern Queensland.

Bichsel, J., (2022). *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research.

Boser, R.S.B.U., (2021). High-leverage opportunities for learning engineering.

Chen, R., (2022). Institutional characteristics and college student dropout risks: A multilevel event history analysis. *Research in Higher education*, *53*, pp.487-505.

Chen, Y. and Upah, S., (2020). Data analytics and STEM student success: The impact of predictive analytics-informed academic advising among undeclared first-year engineering students. *Journal of College Student Retention: Research, Theory & Practice*, *22*(3), pp.497-521.

Daniel, B., (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, *46*(5), pp.904-920.

Del Bonifro, F., Gabbrielli, M., Lisanti, G. and Zingaro, S.P., (2020). Student dropout prediction. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I 21* (pp. 129-140). Springer International Publishing.

Demchenko, Y., (2020), December. From DevOps to DataOps: Cloud based Software Development and Deployment. In *Proc. The International Conference on High Performance Computing and Simulation (HPCS 2020)* (pp. 10-14).

Drachsler, H., Verbert, K., Santos, O.C. and Manouselis, N., (2015). Panorama of recommender systems to support learning. *Recommender systems handbook*, pp.421-451.

Floyd, C., Gulavani, S.S., Du, J., Kim, A.C. and Pappas, J., (2021). A tale of two cities: COVID-19 and the emotional well-being of student-athletes using natural language processing. *Frontiers in Sports and Active Living*, *3*, p.710289.

Houmani, Z., (2021). *Data-driven Management Solution for Microservice-based Deep Learning Applications* (Doctoral dissertation, Université de Lyon).

Ingels, S.J., Pratt, D.J., Wilson, D., Burns, L.J., Currivan, D., Rogers, J.E. and Hubbard-Bednasz, S., (2017). *Education Longitudinal Study of 2002 (ELS): Base-Year to Second Follow-Up Data File Documentation. NCES 2008-347*. ERIC Clearinghouse.

Kovacic, Z., (2020). Early prediction of student success: Mining students' enrolment data.

Kucirkova, N. and Leaton Gray, S., (2023). Beyond Personalization: Embracing Democratic Learning Within Artificially Intelligent Systems. *Educational Theory*, *73*(4), pp.469-489.

Lal, A. and Kumar, G., (2021), July. Intelligent Testing in Software Industry. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 01-06). IEEE.

Li, S., Wang, K., Huo, B., Zhao, X. and Cui, X., (2022). The impact of cross-functional coordination on customer coordination and operational performance: an information processing view. *Industrial management & data systems*, *122*(1), pp.167-193.

Mandinach, E.B. and Gummer, E.S., (2016). *Data literacy for educators: Making it count in teacher preparation and practice*. Teachers College Press.

Marbouti, F., Diefes-Dux, H.A. and Madhavan, K., (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, *103*, pp.1-15.

Martinez, I., Viles, E. and Olaizola, I.G., (2021). Data science methodologies: Current challenges and future approaches. *Big Data Research*, *24*, p.100183.

Masika, R. and Jones, J., (2016). Building student belonging and engagement: Insights into higher education students' experiences of participating and learning together. *Teaching in higher education*, *21*(2), pp.138-150.

McNamara, D.S., Arner, T., Butterfuss, R., Mallick, D.B., Lan, A.S., Roscoe, R.D., Roediger, H.L. and Baraniuk, R.G., (2022). Situating AI (and big data) in the learning sciences: Moving toward large-scale learning sciences. In *Artificial Intelligence in STEM Education* (pp. 289-308). CRC Press.

Nogueira, A.F. and Zenha-Rela, M., (2021). Monitoring a ci/cd workflow using process mining. *SN Computer Science*, *2*(6), p.448.

Osmanbegovic, E. and Suljic, M., (2022). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, *10*(1), pp.3-12.

Pardo, A., (2019). Designing learning analytics experiences. In *Learning analytics: From research to practice* (pp. 15-38). New York, NY: Springer New York.

Park, Y. and Jo, I.H., (2015). Development of the Learning Analytics Dashboard to Support Students' Learning Performance. *J. Univers. Comput. Sci.*, *21*(1), pp.110-133.

Pascarella, E.T. and Terenzini, P.T., (2020). *How College Affects Students: A Third Decade of Research. Volume 2*. Jossey-Bass, An Imprint of Wiley. 10475 Crosspoint Blvd, Indianapolis, IN 46256.

Picciano, A.G., (2022). The evolution of big data and learning analytics in American higher education. *Journal of asynchronous learning networks*, *16*(3), pp.9-20.

Picciano, A.G., (2022). The evolution of big data and learning analytics in American higher education. *Journal of asynchronous learning networks*, *16*(3), pp.9-20.

Poornima, S. and Pushpalatha, M., (2020). A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks*, *1*, pp.76-84.

Rane, N., Choudhary, S. and Rane, J., (2023). Education 4.0 and 5.0: Integrating Artificial Intelligence (AI) for personalized and adaptive learning. *Available at SSRN 4638365*.

Renick, T.M., (2020). Eliminating achievement gaps using data and personalized outreach. *Student Affairs and Services in Higher Education: Global Foundations, Issues, and Best Practices Third Edition*, *23*(2), p.119.

Salmi, J. and Sursock, A., (2018). Access and completion for underserved students: International perspectives. *CIHE Perspectives*, (9), pp.30-32.

Schuh, J.H., (2022). The integrated postsecondary education data system. *New directions for higher education*, *2002*(118), pp.29-38.

Sclater, N., Peasgood, A. and Mullan, J., (2016). Learning analytics in higher education. *London: Jisc. Accessed February*, *8*(2017), p.176.

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P.K., Yuan, X., Nathan, A. and Bhimdiwali, A., (2017). Completing college: A national view of student completion rates–fall 2011 cohort.

Siemens, G. and Long, P., (2021). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), p.30.

Siemens, G., (2023). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), pp.1380-1400.

Slade, S. and Prinsloo, P., (2023). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), pp.1510-1529.

Syed Mustapha, S.M.F.D., (2023). Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods. *Applied System Innovation*, *6*(5), p.86.

Talukder, A.B.M., (2023). Understanding Coordination in Distributed Agile Software Development.

Thomas, L., (2019). Improving student retention in higher education. *Australian Universities Review*, *51*(2), pp.9-18.

Tinto, V., (2015). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago press.

Trowler, V., (2020). Student engagement literature review. *The higher education academy*, *11*(1), pp.1-15.

Tzimas, D. and Demetriadis, S., (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, *69*, pp.1101-1133.

Varlotta, L., (2016). Measuring student success. *The handbook of student affairs administration*, *4*, pp.175-197.

Whitt, E.J., Schuh, J.H., Kinzie, J. and Kuh, G.D., (2023). *Student success in college: Creating conditions that matter*. Hoboken, NJ, USA: Jossey-Bass.

Xu, J., Moon, K.H. and Van Der Schaar, M., (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), pp.742-753.

York, T.T., Gibson, C. and Rankin, S., (2019). Defining and measuring academic success. *Practical assessment, research, and evaluation*, *20*(1), p.5.