

Review of Big Data and Artificial Intelligence in Predicting Higher Education Consumer Preferences

Azad Ali¹ and Dr. Rahul Kushwah²

¹Research Scholar, Department of Management

²Professor, Department of Management

Vikrant University, Gwalior M.P

Abstract: *The rapid expansion of digital technologies has transformed higher education from a traditional service-oriented sector into a data-intensive environment in which institutions increasingly use information about students, applicants, alumni, and prospective learners to understand preferences and improve educational services. Big Data and Artificial Intelligence (AI) provide institutions with opportunities to analyze large and complex datasets generated through admission applications, learning management systems, website interactions, social media, student information systems, surveys, academic records, and digital communication platforms.*

In this context, predicting higher education consumer preferences refers to identifying patterns in the choices, expectations, interests, and behaviors of students and prospective students regarding academic programs, institutions, delivery modes, fees, learning resources, campus services, employability opportunities, and personalized educational experiences. Existing research indicates that Big Data analytics can support administrative decision-making, learning, teaching, student well-being, and institutional planning, while AI can enhance prediction, recommendation, personalization, and automated decision support..

Keywords: Big Data, Consumer Preferences, Predictive Analytics, Student Behaviour, Personalization, Educational Data Mining, Machine Learning, Decision Making.

I. INTRODUCTION

Higher education institutions operate in an increasingly competitive and technology-driven environment where students and prospective learners have multiple choices regarding institutions, academic programs, specialization areas, learning formats, fees, locations, facilities, and career-oriented opportunities. As a result, understanding the preferences of higher education consumers has become an important institutional concern. Traditionally, universities relied on surveys, interviews, admission statistics, feedback forms, and informal observations to understand student expectations. However, digital transformation has created enormous volumes of structured and unstructured information that can provide more detailed insights into consumer behavior.

Big Data enables institutions to collect and analyze information from multiple sources, while Artificial Intelligence allows these data to be transformed into predictions, classifications, recommendations, and behavioral insights. Research on Big Data in higher education shows that institutions increasingly use data generated through student interactions and institutional systems for learning, teaching, administration, and student-related decision-making. Thus, the combination of Big Data and AI has created an opportunity to move from descriptive understanding of student behavior toward predictive approaches that estimate what students may prefer or choose in the future.

BIG DATA IN HIGHER EDUCATION

Big Data in higher education refers to the large-scale collection, processing, and analysis of educational, administrative, behavioral, demographic, and interaction-related information. Higher education institutions generate data through admission applications, registration systems, examination records, learning management systems, online libraries, attendance systems, student feedback, institutional websites, mobile applications, social networking platforms, and digital learning environments. These data may reveal patterns concerning course selection, student engagement, preferred learning resources, academic interests, institutional services, and communication preferences.

A systematic review covering a decade of research found that Big Data and analytics in higher education have been applied mainly to support learning, with additional applications in teaching, administration, and student well-being. The growing availability of such data makes it possible to understand higher education consumers at a more detailed level. For example, analysis of website searches may indicate interest in particular programs, while application and enrollment data can help institutions identify characteristics associated with program selection. Similarly, learning-platform interaction data can provide information about preferred content formats and patterns of student engagement. However, the value of Big Data depends on data accuracy, completeness, integration, and appropriate interpretation.

ARTIFICIAL INTELLIGENCE AND PREDICTIVE ANALYTICS

Artificial Intelligence provides computational methods through which large datasets can be analyzed to identify relationships and generate predictions. Machine learning algorithms such as decision trees, random forests, support vector machines, logistic regression, clustering algorithms, neural networks, and deep learning models can be applied to educational data for classification and prediction. AI-based systems can identify patterns that may not be easily detected through conventional statistical approaches. In higher education, predictive analytics has traditionally been used for areas such as academic performance, retention, dropout risk, student engagement, and recommendation.

Educational Big Data research has specifically identified prediction, recommendation, and evaluation as important applications of educational data analytics. In the context of consumer preferences, AI can extend these applications by predicting the probability that a prospective student will select a particular course, institution, learning mode, specialization, or student service. The increasing sophistication of AI also enables institutions to develop recommendation systems that provide individualized information based on previous behavior and observed preferences.

HIGHER EDUCATION CONSUMER PREFERENCES

Higher education consumers are not limited to enrolled students. The concept may include prospective students, current students, parents or sponsors, international students, adult learners, and other stakeholders involved in educational choices. Their preferences may involve academic quality, course availability, tuition fees, scholarships, institutional reputation, employability, faculty expertise, infrastructure, location, online learning facilities, flexible schedules, student support, campus life, and opportunities for professional development.

Consumer preferences are dynamic and can change according to economic conditions, technological developments, labor-market requirements, personal goals, social influences, and previous educational experiences. AI-based prediction can therefore help institutions identify different consumer segments rather than treating students as a homogeneous population. Recent evidence on AI adoption among post-secondary students indicates that individual, contextual, and technological factors are all associated with students' behavior toward AI technologies, illustrating the multidimensional nature of higher education preferences and technology-related decisions.

MAJOR DATA SOURCES FOR PREFERENCE PREDICTION

Data Source	Examples of Information	Potential Preference Insights
Admission Data	Applications, selected programs, demographic information	Program and institutional preferences
Website Analytics	Page visits, search terms, browsing duration	Areas of academic interest

Learning Management Systems	Logins, content access, activity patterns	Preferred learning resources and formats
Student Information Systems	Enrollment, courses, academic records	Program continuation and selection patterns
Surveys and Feedback	Satisfaction, expectations, opinions	Explicit consumer preferences
Social Media	Comments, interactions, engagement	Perceptions and emerging interests
Mobile Applications	Clicks, searches, notifications	Service and communication preferences
Library Systems	Searches, downloads, resource usage	Information and learning preferences
Financial Data	Fees, scholarships, payment patterns	Price sensitivity and affordability
Career Data	Placement interests, internships, job searches	Career-oriented program preferences

The integration of multiple data sources can provide a more comprehensive understanding of higher education consumers. Nevertheless, data integration should be undertaken carefully because different systems may use different definitions, formats, and levels of accuracy. The ethical use of student-generated data is particularly important because research has reported limited awareness among some students and staff regarding the nature and extent of data collection, analytics, and predictive use.

PREDICTIVE MODELS FOR HIGHER EDUCATION PREFERENCES

Different predictive models can be employed according to the research objective and characteristics of the available dataset. Regression models can estimate the probability of selecting a course or institution, while classification algorithms can categorize students into preference groups. Decision trees and random forests are useful for identifying important variables associated with specific decisions, whereas clustering techniques can identify groups of consumers with similar behavioral patterns.

Neural networks and deep learning can handle complex nonlinear relationships in large datasets. Recommendation systems can combine historical behavior with user characteristics to suggest courses, programs, learning resources, or institutional services. Recent reviews of educational prediction research show continued use of machine learning and deep learning, with tree-based models and neural networks among prominent approaches; explainable AI is also emerging as an important component because institutions need to understand how predictive models reach their outputs.

PERSONALIZATION AND RECOMMENDATION IN HIGHER EDUCATION

One of the major applications of Big Data and AI is personalized educational decision support. Instead of offering identical information to all prospective students, institutions can use predictive models to deliver content according to individual interests and behavioral patterns. For example, a student showing repeated interest in computer science, artificial intelligence, and cybersecurity programs may receive information about relevant courses, scholarships, laboratories, career opportunities, and admission requirements.

Similarly, current students can receive recommendations concerning elective courses, learning materials, academic support, or career services. Research on AI-supported personalized education identifies predictive and diagnostic capabilities as important opportunities for creating student-centered educational experiences. Personalization can potentially improve communication efficiency and student engagement, but it should not become a mechanism that restricts students to predetermined choices. Recommendation systems should therefore provide relevant options while preserving student autonomy.

FACTORS INFLUENCING HIGHER EDUCATION CONSUMER PREFERENCES

Higher education consumer preferences are influenced by demographic, academic, economic, technological, social, and psychological factors. Age, gender, educational background, geographical location, family circumstances, income, career goals, academic achievement, digital literacy, previous institutional experience, and social influence may contribute to decision-making. Institutional factors such as reputation, fees, facilities, course flexibility, faculty quality, employability, scholarships, and technological infrastructure may also influence choices.

AI adoption research in higher education similarly identifies individual, contextual, and technological dimensions as important explanatory categories. Therefore, predictive models should incorporate multiple dimensions instead of depending on a single variable. A multidimensional approach can produce more meaningful consumer segmentation and help institutions understand the complex reasons underlying educational choices.

BENEFITS OF BIG DATA AND AI-BASED PREFERENCE PREDICTION

The integration of Big Data and AI can provide several benefits to higher education institutions. First, predictive analytics can support evidence-based strategic planning by identifying emerging student interests. Second, institutions can improve marketing and communication by providing relevant information to specific consumer groups. Third, prediction can assist course planning by identifying programs with increasing or decreasing demand. Fourth, recommendation systems can improve personalization by connecting students with relevant academic and support opportunities. Fifth, institutions can use historical data to anticipate changes in enrollment demand and allocate resources more efficiently. Sixth, analysis of student feedback and digital interactions can help institutions identify dissatisfaction and improve services. However, these benefits depend on appropriate data governance, valid models, and continuous human oversight rather than on technological implementation alone.

Despite substantial opportunities, Big Data and AI-based prediction raises significant ethical and methodological challenges. Privacy is a central concern because student data may include sensitive personal, academic, behavioral, and financial information. Data collected for educational purposes may potentially be used for other forms of prediction without adequate awareness or consent.

Other concerns include algorithmic bias, inaccurate predictions, data security, lack of transparency, excessive surveillance, and unequal treatment. Research on tertiary education data analytics has emphasized the need for greater attention to student and staff perspectives, ethical issues, policy development, and ethical guidance. Furthermore, predictive models may reproduce existing inequalities if historical data reflect unequal access to educational opportunities. Consequently, institutions should implement data minimization, appropriate consent procedures, security controls, fairness assessment, explainability mechanisms, and human review. AI should support institutional decision-making rather than completely replacing professional judgment or student choice.

COMPARATIVE REVIEW OF MAJOR RESEARCH AREAS

Research Area	Role of Big Data	Role of AI	Expected Outcome	Major Concern
Student Recruitment	Analysis of applicant behavior	Prediction and segmentation	Improved recruitment targeting	Privacy and profiling
Course Selection	Historical enrollment and search data	Recommendation models	Better program matching	Recommendation bias
Learning Preferences	LMS interaction data	Classification and personalization	Personalized learning	Excessive monitoring
Institutional Choice	Application and website data	Predictive classification	Understanding choice patterns	Data quality
Student Services	Feedback and interaction data	Sentiment and prediction models	Improved service delivery	Misinterpretation
Career	Placement and career	Recommendation	Career-oriented	Limited contextual

Preferences	data	systems	guidance	data
Marketing	Digital engagement data	Consumer segmentation	Targeted communication	Manipulative personalization
Retention	Academic and behavioral data	Risk prediction	Early intervention	Labeling and stigma
Policy Planning	Institutional datasets	Forecasting	Strategic planning	Algorithmic dependence

Although substantial research has examined Big Data, learning analytics, educational data mining, and AI adoption, comparatively less attention has been given specifically to the integrated prediction of higher education consumer preferences across the complete decision-making journey. Existing research often concentrates on academic performance, dropout prediction, technology acceptance, or institutional administration rather than examining consumer preferences as a multidimensional construct. The literature also indicates considerable variation in datasets, predictive models, populations, and definitions of AI adoption or educational behavior. A recent meta-analysis of tertiary student AI adoption, for example, identified substantial heterogeneity across studies and highlighted the need to consider AI-specific factors alongside traditional technology-acceptance variables. Therefore, future studies should integrate demographic, behavioral, academic, economic, institutional, and technological variables to develop more comprehensive models of higher education consumer preferences.

FUTURE DIRECTIONS

Future research should focus on developing explainable and ethically responsible AI models that can predict preferences while maintaining student autonomy and privacy. Longitudinal datasets should be used to understand how preferences change over time rather than relying exclusively on cross-sectional data. Researchers should also compare different machine learning and deep learning techniques using standardized evaluation criteria. Explainable AI can help administrators and researchers understand the variables contributing to predictions, which is particularly important when predictions influence admissions, recommendations, marketing, or student support. In addition, future research should examine differences across public and private institutions, disciplines, geographic regions, socioeconomic groups, and online and offline learning environments. Current systematic reviews also suggest that AI adoption patterns vary according to contextual and disciplinary characteristics, supporting the need for more context-sensitive predictive models.

II. CONCLUSION

Big Data and Artificial Intelligence have created significant possibilities for understanding and predicting higher education consumer preferences. Large-scale educational datasets can reveal patterns in student interests, institutional choices, course selection, learning behavior, service expectations, and career objectives, while AI provides methods for transforming these patterns into predictions and recommendations. The literature indicates that Big Data is already being applied across learning, teaching, administration, and student-related activities, while AI is increasingly used for personalization, prediction, recommendation, and decision support. Nevertheless, prediction of consumer preferences should not be treated as a purely technological exercise.

Data quality, privacy, fairness, transparency, explainability, consent, and student autonomy are equally important. The future of AI-driven higher education analytics therefore depends on combining advanced computational methods with responsible data governance and human-centered institutional practices. A comprehensive framework integrating Big Data, AI, consumer behavior, ethical safeguards, and explainable prediction can help higher education institutions understand changing preferences while ensuring that technology remains supportive of informed educational choices.

REFERENCES

1. Braunack-Mayer, A. J., Street, J. M., Tooher, R., Feng, X., & Scharling-Gamba, K. (2020). Student and staff perspectives on the use of Big Data in the tertiary education sector: A scoping review and reflection on the ethical issues. *Review of Educational Research*, 90(6), 788–823.
2. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20, 22.
3. Dimitriadou, S., Athanassopoulos, S., Komis, V., Papadakis, S., & Lavidas, K. (2026). Artificial intelligence in higher education: A systematic review across academic disciplines. *Review of Science, Mathematics and ICT Education*, 20(1).
4. Stojanov, A., & Daniel, B. K. (2024). A decade of research into the application of Big Data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, 29, 5807–5831.
5. Teo, T., Scherer, R., Fung, A. S. K., & Fung, C. S. L. (2026). Factors associated with students' adoption of artificial intelligence technology in tertiary education: A meta-analytic review. *Educational Research Review*, 52, 100804.