Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666

Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

GLOBAL ACADEMIC EXCELLENCE

EVALUATING STUDENT OUTCOMES THROUGH PREDICTIVE MODELING: LESSONS LEARNED IN MALAYSIAN EDUCATION INSTITUTIONS

Muhamad Noorazizi Abd Ghani[1], Istas Fahrurrazi Nusyirwan[2*]

[1] Faculty of Artificial Intelligence, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
Email: muhamadnoorazizi@graduate.utm.my
[2] Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
Email: istaz@utm.my
[*] Corresponding Author

Article Info:

Article history:
Received date: 30.09.2025
Revised date: 10.10.2025
Accepted date: 30.11.2025
Published date: 09.12.2025

To cite this document:

Abd Ghani, M. N., & Nusyirwan, I. F. (2025). Evaluating Student Outcomes Through Predictive Modeling: Lessons Learned In Malaysian Education Institutions. *Journal of Information System and Technology Management, 10* (41), 204-224.

DOI: 10.35631/JISTM.1041013

Abstract:

Evaluating student outcomes through predictive modeling is pivotal for driving timely interventions and enhancing academic success in Malaysian education institutions. This systematic literature review (SLR), conducted following PRISMA guidelines, examined 106 studies published up to 2024 to assess the evolution and efficacy of predictive modeling techniques in this context. The review employed a structured search across Web of Science, Scopus, and Lens.org without publication year restrictions, ensuring comprehensive coverage of the literature. Studies were analyzed for their data collection strategies, and predictive methodologies, ranging from traditional cross-sectional analyses based on academic records to emerging longitudinal designs integrating multimodal data sources such as e-learning logs, surveys, and behavioral metrics. Findings reveal a dominant reliance on academic records, although recent trends indicate a gradual shift toward more sophisticated, multimodal approaches that enhance predictive accuracy. While conventional methods like decision trees and logistic regression remain prevalent, ensemble techniques, deep learning, and hybrid frameworks are increasingly adopted to mitigate challenges such as class imbalance and overfitting. The insights garnered underscore the potential of predictive modeling to inform early intervention strategies and policy decisions, highlighting the need for standardized, multi-institutional, and longitudinal research designs. By addressing these challenges, future research can better support data-driven initiatives that promote student retention, equity, and overall academic excellence in Malaysian educational institutions.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Keywords:

Student Outcomes; Machine Learning; Malaysia; Predictive Modeling; Student Performance; Systematic Literature Review

Introduction

Predicting and enhancing student outcomes has long been a central concern in educational research, especially where timely interventions can reduce dropout and improve institutional effectiveness (Abdul Bujang, Selamat, & Krejcar, 2021). Learning analytics (LA) focuses on collecting and analysing data about learners to understand and optimise learning. A major strand of LA builds predictive models of student success for proactive feedback, typically combining digital trace data from learning management systems with institutional records to identify students at risk and trigger timely interventions (Gašević, Dawson, & Siemens, 2015; Siemens, 2013). In Malaysia, policy initiatives such as the Malaysia Education Blueprint (Ministry of Education, 2013, 2015) promote data-driven decision-making and encourage institutions to deploy predictive analytics to monitor performance and align programme outcomes with labour market needs.

Early Malaysian data mining work in higher education focused mainly on classifying students' success in specific courses using demographic and academic information. Delavari, Beikzadeh, and Phon-Amnuaisuk (2005) used Decision Trees classification to predict success in a core programming course and, through their DM_EDU framework, outlined predictive tasks such as dropout and graduation forecasting. Norwawi, Hibadullah, Abdusalam, and Shuaibu (2009) applied classification algorithms to programming-course performance using Felder-Soloman learning-style profiles with basic background variables. In line with this early work, later Malaysian studies continued to develop predictive models of student success using institutional and assessment data. Aziz, Ismail, and Ahmad (2014) used demographic characteristics and first-semester GPA with classification algorithms, while Suliman, Abidin, Manan, and Razali (2014) modelled pre-university CGPA and subject averages from SPM examination results, gender, and school background. Over time, the methodological landscape has expanded beyond standalone classifiers such as Decision Trees and Naïve Bayes (Abdul Bujang et al., 2021; N. A. A. Rahman, Tan, & Lim, 2017) towards ensemble and deep learning approaches that tackle class imbalance, overfitting, and interpretability and model outcomes beyond binary pass-fail metrics to include performance tiers and longer term academic trajectories (Khamis, Ahmad, Ahmad, & Ahmad, 2022; Teoh, Ho, Dollmat, & Tan, 2023).

From an LA perspective, however, a gap remains between the increasing algorithmic sophistication of these models and their practical integration into institutional strategies. The challenge is no longer just achieving high predictive accuracy, but bridging the divide between offline modelling and actionable interventions. This SLR synthesises 106 Malaysian studies that employ predictive modelling of student outcomes. Anchoring the review in an LA perspective allows us to ask how these studies conceptualise success and how far predictive outputs inform interventions or decision-making. In doing so, the review maps the Malaysian predictive modelling landscape and identifies opportunities for future work to design models and systems that are accurate, usable, interpretable, and aligned with educational goals.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Methodology

This SLR followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The review procedure involved a structured search of selected academic databases, followed by screening and quality assessment to ensure the inclusion of methodologically sound and contextually relevant studies.

Search Strategy

The primary search was performed in three academic databases: Web of Science (WoS), Scopus, and Lens.org. These databases were chosen due to their broad coverage of multidisciplinary research and comprehensive indexing of scholarly publications. The search employed a combination of keywords and Boolean operators to capture studies related to predictive modelling in Malaysian education institutions. Specifically, the following search terms were used:
(predict* OR forecast* OR projection* OR model* OR analytic* OR estimate*)
AND
(academic* OR educational* OR "higher education" OR "academic achievement" OR "academic performance")
AND
(student* OR learner* OR "student performance" OR "student success" OR "student retention")

No restriction was placed on publication year; all articles indexed up to 2024 were considered. The initial search was conducted on August 12, 2024, followed by a supplementary search on December 12, 2024, to capture newly published or updated content. The initial search identified 921,968 records: 140,292 from WoS, 159,091 from Scopus, and 622,585 from Lens.org. See Figure 1.

Inclusion and Exclusion Criteria

Inclusion Criteria:

- Country: Studies conducted in Malaysia.
- Document Type: Research articles, conference proceedings articles, and book chapters.
- Language: English, Malay, or Indonesian.

Exclusion Criteria:

- Studies focusing on countries other than Malaysia.
- Document types other than research articles, conference proceedings, and book chapters.
- Languages other than English, Malay, or Indonesian.

After applying these criteria, the number of articles was reduced to 10,460: WoS (2,778), Scopus (3,337), and Lens.org (4,345).

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Screening and Selection Process
The identified studies underwent a multi-phase screening process:

1. **Deduplication**: 2,342 duplicates were removed, leaving 8,118 unique records.
2. **Title and Abstract Screening**: The remaining 8,118 records were screened based on their titles and abstracts. Studies not aligning with the inclusion criteria were discarded, particularly those not focusing on predictive approaches, the Malaysian context, student-related outcomes, and acceptable publication types. This step removed 7,276 articles.
3. **Full-Text Availability Check**: Among the remaining records, 36 articles were excluded as full-text versions were inaccessible.

Following these steps, 806 articles were retained for full-text review.

Full-Text Review and Quality Assessment
Each of the 806 full-text articles was assessed for methodological rigour and relevance. Articles were examined for the presence of predictive modelling aimed at student outcomes in Malaysian education institutions. Studies failing these quality assessments were removed at this stage:

- 530 studies did not involve predictive modelling and were thus excluded.
- 161 studies used public or non-Malaysian origin datasets, contradicting the inclusion requirements.
- 11 studies did not focus on students or student-related outcomes.

After this quality assessment, 104 articles remained. A subsequent manual search of reference lists and relevant literature led to the inclusion of an additional two articles, resulting in a total of 106 studies that met all eligibility and quality criteria.

The final set of 106 articles represents the corpus of literature selected for in-depth review and analysis. These studies collectively form the evidence base for the SLR, ensuring a methodologically sound, contextually focused, and comprehensive understanding of predictive modelling in Malaysian education institutions.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

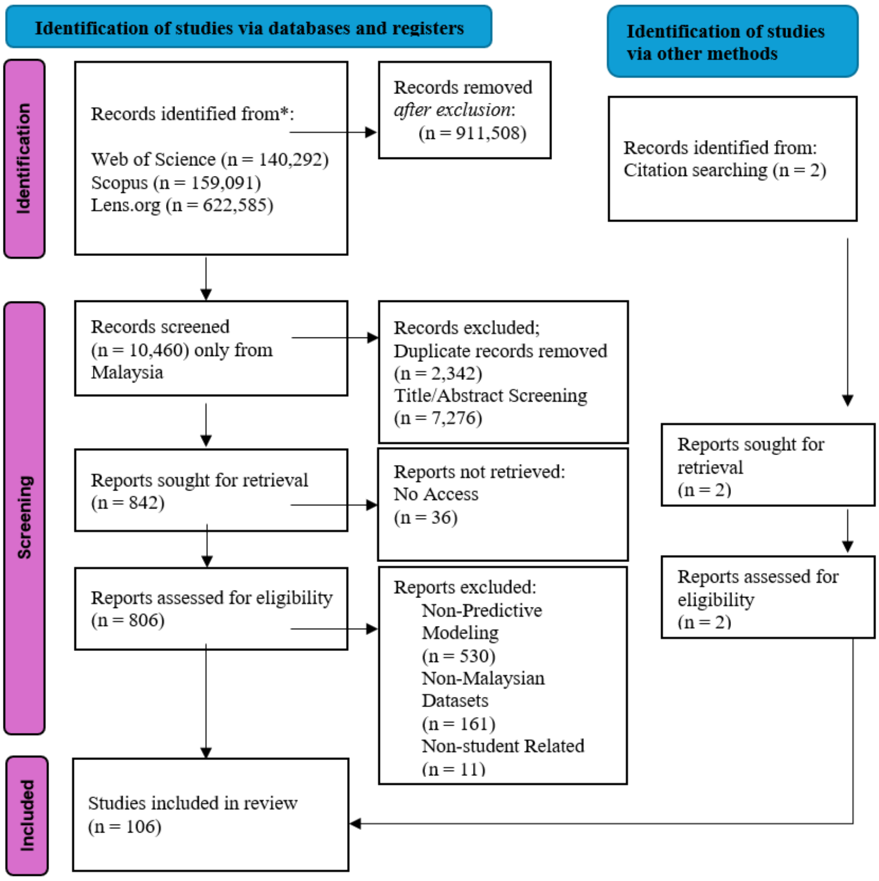

Figure 1: Flowchart of Article Selection

Results

This review synthesises key studies on predictive modelling in Malaysian education. These studies provide a foundation for mapping how methods, data practices, and modelling choices have evolved across the local research landscape.

Methodological and Data Scope

This section synthesises the data usage in Malaysian educational institutions and student outcomes prediction research. Drawing upon 106 research studies that utilise Malaysian samples or datasets.

Data Collection Methods

Academic records remain the primary data source across almost all studies, offering key indicators such as Cumulative Grade Point Average (CGPA), semester grades, entry qualifications (e.g., SPM, MUET), and demographic and socioeconomic details. From the earlier period (2014 and prior), in the majority of the reviewed studies (Abidin, Setu, Yong, Foong, & Ahmad, 2008; Affendey, Paris, Mustapha, Sulaiman, & Muda, 2010; Arsad et al., 2012, 2013b; Azmi & Paris, 2011; Delavari et al., 2005; Hasan, Adam, Mustapha, & Bakar, 2012; Hoe et al., 2013; Lye et al., 2010; Noor, Kadirgama, Rahman, Bakar, & Ibrahim, 2010; Rani & Embong, 2013; Suliman et al., 2014), academic records from institutional databases

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
(e.g., student demographics, grades, CGPA) serve as the primary data collection source. A smaller subset relies on survey instruments (Norwawi et al., 2009; Sembiring, Zarlis, Hartama, S, & Wani, 2011; Siraj et al., 2006; Wook et al., 2009), typically focusing on learning styles, personality traits, or graduate outcomes, and a few studies incorporate additional data such as personal records or interaction logs (Sembiring et al., 2011; Ting, Cheah, & Ho, 2013). Most studies concentrate on institutional academic data as the backbone for predictive modelling, occasionally enriched by supplementary survey or engagement data.

Between 2015 and 2019, the dominant data collection method remains academic records sourced from university databases and institutional systems (A. A. Aziz, Ismail, Ahmad, & Hassan, 2015; F. Aziz et al., 2015; Ghani, Cob, Drus, & Sulaiman, 2019; Makhtar, Nawang, & Shamsuddin, 2017; Razak, Omar, & Ahmad, 2018). These typically include demographic information, prior academic performance (such as CGPA), and additional institutional data (e.g., entry qualifications or sponsorship details). Several studies supplement these records with surveys, most notably Malaysia's Ministry of Higher Education Tracer Study (Z. Othman et al., 2018; N. A. A. Rahman et al., 2017; N. A. B. A. Rahman, Tan, & Lim, 2017; Siraj & Bakar, 2019), which captures graduate employment outcomes and personal profiles. A few researchers incorporate e-learning log data or online activity (Hassan, Anuar, & Ahmad, 2019; Mohamad, Ahmad, & Jawawi, 2018; Shukor et al., 2015), while others gather pre-and post-test assessments to measure learning gains (Shukor et al., 2015). Although no significant anomalies emerge, some studies focus on unique attributes such as co-curricular involvement (M. T. R. A. Aziz & Yusof, 2016) or specific quality-of-life surveys (Raihana & Nabilah, 2018). Overall, the main pattern across these studies is using institutional academic data as a primary source, occasionally enriched by survey instruments, test results, or system log files.

Across the more recent studies between 2020 and 2024, academic records remain the predominant data source, typically including demographic details, grades, CGPA, and program-related attributes (W. D. Ahmad & Bakar, 2020; Kumaran, Yusuf, Othman, & Yunianta, 2020; Sani et al., 2020; Wan Yaacob et al., 2020). Many authors supplement institutional records with survey data, such as Malaysia's graduate tracer studies or custom questionnaires that capture variables like learning styles, self-efficacy, and study habits (Ghazvini, Sharef, & Sidi, 2024; Looi, Song, Lim, & Looi, 2024; N. H. A. Rahman, Sulaiman, & Ramli, 2023). A notable trend is the increasing use of system logs, for example, e-learning, microlearning, or learning management system (LMS) platforms, to track student activity and engagement (Ali, Thomas, & Nair, 2021; Hamzah, Yusoff, Ismail, & Ismail, 2021; Hassan, Ahmad, & Anuar, 2020). A few studies incorporate psychometric tests (N. A. Zakaria, Ahmad, Awang, & Safar, 2021), interviews with domain experts (Lumius & Asli, 2021), or library activity data (Ali et al., 2021), indicating a broader range of behavioural and contextual factors being explored. While there are no significant anomalies, some studies rely solely on surveys for self-reported measures (Borhani & Wong, 2023), and a few merge multiple data streams (e.g., Kaggle datasets, offline exam records) (Rahim & Buniyamin, 2023), highlighting an ongoing shift toward more affluent, multimodal data collection.

Dataset and Sample Size

Many studies centre on single-institution datasets, often from a specific faculty or department, while multi-institution studies are less frequent. 2014 and prior, most studies utilized a single dataset (Abidin et al., 2008; Affendey et al., 2010; Arsad et al., 2012; Azmi & Paris, 2011; Lye et al., 2010; Noor et al., 2010; Rani & Embong, 2013; Sapaat, Mustapha, Ahmad, Chamili, &

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Muhamad, 2011; Siraj et al., 2006; Suliman et al., 2014), and several integrated multiple sources were also utilized by some studies (Delavari et al., 2005; Hoe et al., 2013; Sembiring et al., 2011). Sample sizes varied greatly, from as few as 54 students (Ting et al., 2013) to over 22,000 (Hasan et al., 2012), with most datasets focusing on student academic records and demographic information. While nearly all studies explicitly stated their sample sizes, one omitted the precise figure (A. A. Aziz, Ismail, & Ahmad, 2013). Despite their methodological differences, the recurring pattern emphasises predicting or classifying student outcomes through data mining or statistical models, underscoring the value of small-scale and large-scale educational data to uncover performance indicators.

Summarising studies from 2015 to 2019 reveals that most studies used either a single dataset or integrated multiple datasets, commonly combining student records, academic logs, or tracer study data (Ahmad Tarmizi, Mutalib, Abdul Hamid, Abdul-Rahman, & Md Ab Malik, 2019; Chuan et al., 2017; Shariff, Rodzi, Rahman, Zahari, & Deni, 2016). Sample sizes varied greatly, from as low as 20 participants (Shukor et al., 2015) to over 119,000 (A'rifian, Daud, Romzi, & Shahri, 2019), with several studies exceeding tens of thousands of records (Z. Othman et al., 2018; Siraj, 2016). Anomalies typically included missing or incomplete data leading to smaller analysed samples (Shariff et al., 2016) or the absence of explicit sample size counts in a few studies (Mohamad et al., 2018; F. Zakaria, Kar, Abdullah, & Ismail, 2019).

From 2020 to 2024, most studies on predicting student performance or related outcomes relied on single datasets from academic records, surveys, or blended sources (Borhani & Wong, 2023; Sani et al., 2020). Several studies integrated multiple datasets, often combining institutional databases with log files or surveys, to gain broader insights (W. D. Ahmad & Bakar, 2020; Chan & Ng, 2024; Lestari et al., 2024). Sample sizes exhibited a stark range, from fewer than 60 students (N. Ahmad, Hassan, Jaafar, & Enzai, 2021) to over 109,000 records (Hashim, Lim, Jafar, Shanmugam, & Bukhari, 2024), underscoring small-scale and large-scale analyses. A recurring pattern involved data-cleaning steps that often reduced initial numbers (M. Z. M. Sabri et al., 2023), while some studies did not explicitly disclose final sample counts (Samsudin et al., 2021).

Modelling Inputs and Outputs

This section examines how the studies identified in our SLR define their predictors and target variables.

Predictors and Target Variables

This section explores the dual dimensions of predictors and target variables in student outcomes studies. It begins by examining the vast array of predictors leveraged to predict outcomes, from demographic and academic factors to socio-behavioural and unconventional predictors. Next, it discusses the diversity of target variables, ranging from binary and multi-class classifications to continuous regression outputs like CGPA predictions. Finally, the section highlights the best-performing predictors identified in the studies, emphasising their evolving role in enhancing the accuracy and relevance of predictive models.

Predictors

Studies from 2014 and prior reveal that a wide range of demographic, academic, personality, and skill-based factors have been consistently used as predictors in student outcomes studies (Abidin et al., 2008; Affendey et al., 2010; Arsad et al., 2012; Azmi & Paris, 2011; Delavari et

Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
al., 2005; Hasan et al., 2012; Hoe et al., 2013; Lye et al., 2010; Noor et al., 2010; Norwawi et al., 2009; Rani & Embong, 2013; Sapaat et al., 2011; Sembiring et al., 2011; Siraj et al., 2006; Suliman et al., 2014; Ting et al., 2013; Wook et al., 2009). Recurring themes include prior academic achievement (e.g., secondary exam results, GPA), demographic predictors (e.g., gender, age, race, family income), and personal predictors (e.g., learning style, personality traits, language proficiency). Several studies also highlight course-specific grades, sponsorship status, institution type, and specialised features, such as the lecturer's background (Delavari et al., 2005) and students' engagement in hypothesis formulation (Ting et al., 2013). Although most predictors focus on grades and demographic variables, some incorporate socio-behavioural factors like family support and personal motivation (Sembiring et al., 2011), as well as job status or reasons for unemployment (Sapaat et al., 2011), demonstrating the breadth of data used to predict student outcomes.

The following period of the year (2015-2019) indicates that studies on student outcomes broadly rely on demographic (e.g., age, gender, marital status), academic (e.g., subject-specific grades, GPA/CGPA, entry qualification), and socio-economic (e.g., family income, sponsorship) predictors (F. Ahmad et al., 2015; A. A. Aziz et al., 2015; Sangodiah et al., 2015). More recent studies integrate behavioural predictors, such as login frequency, resource viewing, and forum interactions from e-learning platforms (Hassan et al., 2019; Shukor et al., 2015), and incorporate lifestyle or psychosocial predictors, including medical status, part-time job status, or quality-of-life scores (Chuan et al., 2017; Raihana & Nabilah, 2018). Several studies also include less conventional predictors, such as co-curricular involvement, number of supervisors, or campus visit programs, reflecting a growing interest in holistic indicators of student success (M. T. R. A. Aziz & Yusof, 2016; Chin et al., 2019). Notable anomalies include unconventional features, like daily diet, timestamped MOOC logs, and detailed personal identification fields (Chuan et al., 2017; Ghani et al., 2019; Mohamad et al., 2018), suggesting ongoing experimentation with multifaceted data to refine predictive models.

Studies from 2020 to 2024 indicate that most studies continue to use a broad mix of demographic (e.g., age, gender, marital status), academic (e.g., GPA/CGPA, course grades, prior qualifications), and socio-behavioural (e.g., learning engagement, internet usage, family background) predictors (W. D. Ahmad & Bakar, 2020; Kumaran et al., 2020; Sani et al., 2020). Many studies have broadened their scope by integrating online or blended learning predictors, such as LMS activity, assignment submission frequency, and video engagement (Borhani & Wong, 2023; Hassan et al., 2020; Tan et al., 2022). Others focus on specialised predictors like Program Learning Outcomes (PLOs), financial aid eligibility, or behavioural engagement (Hamzah et al., 2021; Ismail, Razak, Noor, & Aziz, 2024; W. N. A. W. Othman, Abdullah, & Romli, 2020). Anomalies include studies examining very narrow or unique features (e.g., age as the sole predictor or multiple intelligences profiles) and those capturing particular demographic features, such as the number of *surahs* (chapters) memorised or orphan status (Hashim et al., 2024; N. A. Zakaria et al., 2021; Zulfikri, Shaharudin, Rajak, & Ibrahim, 2021). Across these studies, the ongoing trend is toward more granular, multifaceted predictors, combining traditional academic records with in-depth, real-time behavioural data to enhance predictive accuracy.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Target Variable

Between 2005 and 2014, the examined studies predominantly focused on predicting or classifying student outcomes through predictive modelling approaches. The target variables range from binary indicators (e.g., pass/fail) to multi-class grade distinctions and continuous regression-based predictions of CGPA (Abidin et al., 2008; Arsad et al., 2012; Delavari et al., 2005; Noor et al., 2010; Siraj et al., 2006). Common categories of target variables include binary classifications (e.g., successful/unsuccessful), multi-class classifications (e.g., letter grades or class standings), and regression outputs (e.g., CGPA). Notable variations arise in the specific outcome labels (e.g., employment status, engagement levels) and the granularity of academic performance categories (e.g., “Excellent” to “Poor,” or “First Class” to “Third Class”).

From 2015 to 2019, the range of target variables broadened to include more distinctive categories of student outcomes, such as detailed academic performance levels (e.g., “Excellent,” “Good,” “Average,” “Poor”), employment status (e.g., “Employed,” “Further Study”), and student retention or attrition (A. A. Aziz et al., 2015; Mohamad et al., 2018; Sangodiah et al., 2015; Shariff et al., 2016). While multi-class classification remains typical for differentiating grades and performance tiers, binary classification also frequently occurs, particularly for target variables such as “Graduate on Time” (GOT) versus non-GOT, “Employed” versus “Not Employed,” and enrollment decisions like “Accept” versus “Reject” (A’rifian et al., 2019; Basheer, Mutalib, Hamid, Abdul-Rahman, & Malik, 2019; Chuan et al., 2017). A subset of studies employs regression to predict CGPA at graduation, reflecting continued interest in continuous numeric prediction (Razak et al., 2018; Siraj, 2016). Notable variations include highly granular multi-class groupings, exemplified by Shukor et al.’s (2015) detailed “Outstanding” to “Poor” categories, highlighting the trend toward increasingly specific student outcomes. Researchers have expanded their exploration beyond traditional pass-fail metrics, with anomalies primarily centred around specialised classification schemes and unique decision cutoffs.

From 2020 to 2024, studies continue to expand beyond traditional target variables, such as binary pass/fail classifications and multi-class final grade distributions, by incorporating broader student outcomes like dropout risk (Sani et al., 2020; Tan et al., 2022), employability (Haque, Quek, Ting, Goh, & Hasan, 2024; W. N. A. W. Othman et al., 2020), scholarship eligibility (W. D. Ahmad & Bakar, 2020), and financial aid decisions (Ismail et al., 2024). Regression-based studies remain prevalent, particularly for predicting numeric outcomes such as CGPA and course grades over time, reflecting sustained interest in continuous performance prediction (Fuad et al., 2021; Samsudin et al., 2022; Zulfikri et al., 2021). Notable anomalies involve specialised target variables, including 'days from graduation to first employment' or contextual impacts such as the effects of COVID-19 on student performance, underscoring the evolving scope of predictive modelling in Malaysian education institutions (Fuad et al., 2021; W. N. A. W. Othman & Abdullah, 2020).

Machine Learning Approaches to Predict Outcomes

This section synthesises the machine learning approaches used to predict, classify, or forecast student outcomes in Malaysian education institutions. These outcomes include academic performance, dropout risk, and graduation timelines.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

Classification Algorithms

Classification has long been the most popular approach in Malaysian educational research for predicting student outcomes. Early studies (2014 and prior) often relied on Decision Trees, such as C4.5 and CART, to categorise students into at-risk or high-performing groups based on attendance, grades, and demographics (Delavari et al., 2005; Siraj et al., 2006). From 2015 to 2019, classification expanded to more diverse methods, including SVM for dropout detection (Sangodiah et al., 2015) and Logistic Regression for predicting on-time PhD completions (Shariff et al., 2016). This period also saw Bayesian methods, particularly Naïve Bayes, gaining attention for their speed and simplicity when classifying student employability (Aziz et al., 2015). Between 2020 and 2024, classification frameworks recently integrated neural networks and hybrid models for multi-class labelling and early detection of performance issues (Bujang et al., 2021; Lumius & Asli, 2021). Across these eras, classification studies consistently emphasise predictive accuracy, student engagement signals, and practical applications in early-warning systems.

Regression Algorithms

Regression-based methods focus on modelling continuous academic indicators like GPA or time-to-graduation. Before 2015, researchers used Multiple Linear Regression (MLR) to link entry qualifications to final CGPA, as seen in studies modelling engineering students' performance (Noor et al., 2010). From 2015 to 2019, regression approaches became more sophisticated, incorporating Radial Basis Function Networks to handle intricate relationships in pre-university performance (Suliman et al., 2014). From 2020 onward, logistic regression, often counted as a classification method, employed quasi-regression when exploring nonlinear terms (Chan & Ng, 2024). Deep learning regressors like LSTM variants have been introduced to track sequential data (Fuad et al., 2021) and address pandemic-era challenges.

Clustering Algorithms

Clustering techniques offer a way to identify hidden groupings among learners, revealing patterns that might not be apparent through supervised methods. In earlier studies (before 2015), kernel k-means and standard k-means were used to group students by psychometric or learning style factors (Sembiring et al., 2011). During 2015-2019, clustering gained traction in multi-step frameworks: for instance, some studies clustered students by learning behaviour before applying classification, thereby refining prediction accuracy and interpretability (Nawang et al., 2018). In recent years (2020-2024), unsupervised approaches have diversified to include hierarchical clustering, CobWeb, and Expectation-Maximization, targeting dimensions like internet usage behaviour and engagement data (Khamis et al., 2022). These analyses help institutions spot at-risk subgroups, tailor interventions to each cluster's needs, and refine admission or tutoring strategies without requiring prior labels for student outcomes or performance classes.

Ensemble Methods and Hybrid Approaches

Ensemble learning combines multiple base learners, such as Decision Trees, Naïve Bayes, or neural networks, to enhance prediction robustness. Before 2015, studies showcased how bagging or boosting with tree-based methods improved accuracy for graduate employability classification (Sapaat et al., 2011). Between 2015 and 2019, ensembles like Random Forest (RF) and AdaBoost rose in popularity, benefiting from greater computing power and more extensive datasets (Hassan et al., 2019). As of 2020 onward, Gradient Boosting variants (XGBoost, LightGBM, CatBoost) have garnered considerable attention for tackling

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
imbalanced classes or high-dimensional features (Hassan et al., 2020; Law, Ting, Ng, et al., 2024). Researchers also explored stacking multiple classifiers by mixing SVM, Artificial Neural Networks (ANN), and RF to address multi-class performance prediction (Bujang et al., 2021). These hybrid frameworks underscore the growing emphasis on performance generalisation, as they systematically blend the strengths of individual models while mitigating weaknesses, thereby boosting overall predictive power.

Discussion

This review of 106 studies on predictive modelling in Malaysian education institutions reveals methodologically, the field has progressed from small, cross-sectional case studies that rely almost exclusively on academic records to more recent work that integrates e-learning traces and behavioural indicators. At a high level, however, the dominant pattern remains one of single-institution, one-off modelling exercises rather than sustained LA programmes embedded in institutional practice.

A first key lesson is that technical sophistication has grown faster than educational integration. Ensemble and deep learning methods are now routinely reported as outperforming traditional classifiers on accuracy scores, particularly in imbalanced settings. Yet these gains are often incremental, rarely benchmarked against strong baselines, and seldom linked to concrete improvements in retention or graduate outcomes. In many studies, the best model is chosen based on marginal performance differences on historical data, with limited reflection on interpretability or alignment with institutional decision cycles.

A second lesson concerns the persistent gap between prediction and action. Across the corpus, most articles stop at reporting predictive performance and identifying important features such as prior grades and engagement measures. Only a small minority of studies move beyond offline modelling to design dashboards, early-warning systems, or decision-support tools such as OBE-related visual analytics prototypes. Even within this subset, evaluation typically focuses on system accuracy or user satisfaction rather than on whether the intervention changed teaching practices, reduced dropout. In other words, the frequently implied pipeline, build a predictive model, deploy it, and student outcomes will improve, remains more an assumption than an empirically tested pathway.

Third, several unresolved challenges cut across data and methods, or even implementation. Data quality issues like missingness, inconsistent cohorts, small samples and the predominance of single-institution datasets continue to limit generalisability. Methodologically, many studies still rely on basic preprocessing and validation strategies, making it difficult to judge robustness under deployment conditions or across cohorts. From a socio-technical perspective, questions of governance and ethics are seldom highlighted, even though they are likely to determine whether predictive tools are trusted and used in practice.

Taken together, these findings suggest that Malaysian predictive modelling research has laid a strong technical and contextual foundation but has yet to fully realise its promise for students and institutions. Future work needs to shift from asking which model predicts best to asking under what conditions, and through which intervention mechanisms, do predictive insights meaningfully improve curriculum and instructional approaches, retention, enhanced monitoring and graduate trajectories. Addressing this will require theory-informed and

Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

longitudinal studies that explicitly design, implement, and rigorously evaluate interventions built on predictive models, rather than treating prediction as an end in itself.

Conclusion

This review has synthesised 106 Malaysian studies on predictive modelling of student outcomes, mapping an active but uneven research landscape. Collectively, these studies demonstrate that data-driven approaches can identify at-risk students and illuminate patterns in progression and achievement. At the same time, the field remains dominated by single-institution and cross-sectional designs, with relatively few theory-informed, institutionally embedded interventions.

Looking ahead, a key priority is to move from predicting risk to changing outcomes. Future work should design and evaluate longitudinal, intervention-centred studies in which predictive models trigger concrete actions, such as targeted advising or remedial support, and track their impact on retention and success over multiple cohorts. Experimental or quasi-experimental designs, grounded in student integration and learning analytics frameworks, are needed to establish when and for whom predictive systems make a meaningful difference.

A second agenda concerns data governance and methodological standards. As institutions integrate richer data sources (academic, behavioural, socio-economic, and learning-platform traces), there is an urgent need for context-sensitive frameworks for ethical data use in Malaysian educational settings. These should address consent, transparency, algorithmic bias, and governance structures, alongside clearer reporting of preprocessing, validation, and computational environments to support reproducibility and cross-study comparison.

Third, future research should examine scalability and transferability across diverse institutions, including public and private universities or TVET providers. Multi-institutional studies that test whether models travel across contexts, or how they must be adapted, would clarify the balance between local tailoring and generalisable design. This includes exploring shared infrastructure and co-designed dashboards that can be integrated into everyday academic and administrative workflows.

Advancing along these lines would shift Malaysian predictive modelling from isolated technical experiments toward robust socio-technical systems that are theoretically grounded and ethically governed, which can improve student trajectories and institutional decision-making.

Acknowledgements

The authors would like to acknowledge and extend special gratitude to Mohd Faiz Mustafa from the Faculty of Pharmacy, Universiti Teknologi Mara, Puncak Alam, Selangor, Malaysia, who reviewed and suggested some improvements for this paper.

References

A'rifian, N. I. N. B., Daud, N. S. A. B. M., Romzi, A. F. B. M., & Shahri, N. H. N. B. M. (2019). A Comparative Study on Graduates' Employment in Malaysia by using Data Mining. *Journal of Physics: Conference Series*, *1366*(1). https://doi.org/10.1088/1742-6596/1366/1/012120

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Abdul Bujang, S. D., Selamat, A., & Krejcar, O. (2021). A Predictive Analytics Model for Students Grade Prediction by Supervised Machine Learning. *IOP Conference Series: Materials Science and Engineering*, *1051*(1), 012005. https://doi.org/10.1088/1757-899x/1051/1/012005

Abidin, A. I. Z., Setu, I. A., Yong, S. P., Foong, O. M., & Ahmad, J. (2008). Classifying Student Academic Performance: A Hybrid Approach. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, *I*, 19–21. Retrieved from http://eprints.utp.edu.my/1182/

Adnan, N. I. M., Mohamed, A. S. T., Azami, M. F. A. M., & Razali, F. A. (2021). Multiple Linear Regression of Asia Pacific University Malaysia Students' Performance in Statistics and Mathematics Course Using R Software. *AIP Conference Proceedings*, *2355*. https://doi.org/10.1063/5.0053195

Affendey, L. S., Paris, I. H. M., Mustapha, N., Sulaiman, M. N., & Muda, Z. (2010). Ranking of Influencing Factors in Predicting Students' Academic Performance. *Information Technology Journal*, *9*(4), 832–837. https://doi.org/10.3923/itj.2010.832.837

Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The Prediction of Students' Academic Performance using Classification Data Mining Techniques. *Applied Mathematical Sciences*, *9*(3), 6415–6426. https://doi.org/10.12988/ams.2015.53289

Ahmad, N., Hassan, N., Jaafar, H., & Enzai, N. I. M. (2021). Students' Performance Prediction using Artificial Neural Network. *IOP Conference Series: Materials Science and Engineering*, *1176*(1), 012020. https://doi.org/10.1088/1757-899x/1176/1/012020

Ahmad Tarmizi, S. S., Mutalib, S., Abdul Hamid, N. H., Abdul-Rahman, S., & Md Ab Malik, A. (2019). A Case Study on Student Attrition Prediction in Higher Education Using Data Mining Techniques. *Communications in Computer and Information Science*, *1100*, 181–192. https://doi.org/10.1007/978-981-15-0399-3_15

Ahmad, W. D., & Bakar, A. A. (2020). Ensemble Machine Learning Model for Higher Learning Scholarship Award Decisions. *International Journal of Advanced Computer Science and Applications*, *11*(5), 303–312. https://doi.org/10.14569/IJACSA.2020.0110540

Ali, A. M., Thomas, J. J., & Nair, G. (2021). Academic and Uncertainty Attributes in Predicting Student Performance. In *Intelligent Computing and Optimization: Proceedings of the 3rd International Conference on Intelligent Computing and Optimization 2020 (ICO 2020)* (pp. 838–847). https://doi.org/10.1007/978-3-030-68154-8_72

Arsad, P. M., Buniyamin, N., & Manan, J.-L. A. (2013a). The Impact of English on Students Performance based on Neural Network Prediction: A Malaysian Case Study. In *Recent Advances in Modern Educational Technologies* (pp. 81–86).

Arsad, P. M., Buniyamin, N., & Manan, J. A. (2012). Neural Network Model to Predict Electrical Students' Academic Performance. *ICEED 2012 - 2012 4th International Congress on Engineering Education - Improving Engineering Education: Towards Sustainable Development*, 1–5. https://doi.org/10.1109/ICEED.2012.6779270

Arsad, P. M., Buniyamin, N., & Manan, J. A. (2013b). A Neural Network Students' Performance Prediction Model (NNSPPM). *2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2013*, (July 2006), 1–5. https://doi.org/10.1109/ICSIMA.2013.6717966

Arsad, P. M., Buniyamin, N., & Manan, J. A. (2014). Neural Network and Linear Regression Methods for Prediction of Students' Academic Achievement. *2014 IEEE Global Engineering Education Conference (EDUCON)*, (April), 916–921. IEEE. https://doi.org/10.1109/EDUCON.2014.6826206

Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Aziz, A. A., Ismail, N. H., & Ahmad, F. (2013). Mining Students' Academic Performance. *Journal of Theoretical and Applied Information Technology*, *53*(3), 485–495.

Aziz, A. A., Ismail, N. H., & Ahmad, F. (2014). First Semester Computer Science Students' Academic Performances Analysis by Using Data Mining Classification Algorithms. *Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014)*, (September), 100–109. WorldConferences.net. Retrieved from http://worldconferences.net

Aziz, A. A., Ismail, N. H., Ahmad, F., & Hassan, H. (2015). A Framework for Students' Academic Performance Analysis using Naïve Bayes Classifier. *Jurnal Teknologi*, *75*(3), 13–19. https://doi.org/10.11113/jt.v75.5037

Aziz, F., Jusoh, A. W., & Abu, M. S. (2015). A Comparison of Student Academic Achievement Using Decision Trees Techniques: Reflection from University Malaysia Perlis. *AIP Conference Proceedings*, *1660*(April), 050034. https://doi.org/10.1063/1.4915667

Aziz, M. T. R. A., & Yusof, Y. (2016). Graduates Employment Classification using Data Mining Approach. *AIP Conference Proceedings*, *1761*, 020002. https://doi.org/10.1063/1.4960842

Azmi, M. S. B. M., & Paris, I. H. B. M. (2011). Academic Performance Prediction Based on Voting Technique. *2011 IEEE 3rd International Conference on Communication Software and Networks, ICCSN 2011*, 24–27. https://doi.org/10.1109/ICCSN.2011.6014841

Baashar, Y., Hamed, Y., Alkawsi, G., Capretz, L. F., Alhussian, H., Alwadain, A., & Al-amri, R. (2022). Evaluation of Postgraduate Academic Performance using Artificial Intelligence Models. *Alexandria Engineering Journal*, *61*(12), 9867–9878. https://doi.org/10.1016/j.aej.2022.03.021

Basheer, M. Y. I., Mutalib, S., Hamid, N. H. A., Abdul-Rahman, S., & Malik, A. M. A. (2019). Predictive Analytics of University Student Intake using Supervised Methods. *IAES International Journal of Artificial Intelligence*, *8*(4), 367–374. https://doi.org/10.11591/ijai.v8.i4.pp367-374

Borhani, K., & Wong, R. T. K. (2023). An Artificial Neural Network for Exploring the Relationship between Learning Activities and Students' Performance. *Decision Analytics Journal*, *9*(August), 100332. https://doi.org/10.1016/j.dajour.2023.100332

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, *9*, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

Chan, L. G., & Ng, Q. Y. (2024). Using Learning Analytics in Higher Education: Assessing Students' Learning Experience in an Actuarial Science Course. *STEM Education*, *4*(2), 151–164. https://doi.org/10.3934/steme.2024010

Chin, W. Y., Ch'ng, C. K., & Jamil, J. M. (2019). A Study of Graduate on Time (GOT) for Ph.D Students using Decision Tree Model. *AIP Conference Proceedings*, *2138*(August). https://doi.org/10.1063/1.5121085

Chuan, Y. Y., Husain, W., & Shahiri, A. M. (2017). An Exploratory Study on Students' Performance Classification Using Hybrid of Decision Tree and Naïve Bayes Approaches. *Advances in Intelligent Systems and Computing*, *538 AISC*, 142–152. https://doi.org/10.1007/978-3-319-49073-1_17

Delavari, N., Beikzadeh, M. R., & Phon-Amnuaisuk, S. (2005). Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. *2005 6th*

Journal of Information System and Technology Management JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
International Conference on Information Technology Based Higher Education and Training, *2005*, F4B-1-F4B-6. IEEE. https://doi.org/10.1109/ITHET.2005.1560303

Fuad, M. F. M., Shaharudin, S. M., Ismail, S., Samsudin, N. A. M., & Zulfikri, M. F. (2021). Comparison of Singular Spectrum Analysis Forecasting Algorithms for Student's Academic Performance during COVID-19 Outbreak. *International Journal of Advanced Technology and Engineering Exploration*, *8*(74), 178–189. https://doi.org/10.19101/IJATEE.2020.S1762138

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Ghani, N. L. A., Cob, Z. C., Drus, S. M., & Sulaiman, H. (2019). Student Enrolment Prediction Model in Higher Education Institution: A Data Mining Approach. In *Lecture Notes in Electrical Engineering* (Vol. 565). Springer International Publishing. https://doi.org/10.1007/978-3-030-20717-5_6

Ghazvini, A., Sharef, N. M., & Sidi, F. B. (2024). Prediction of Course Grades in Computer Science Higher Education Program via a Combination of Loss Functions in LSTM Model. *IEEE Access*, *12*(January), 30220–30241. https://doi.org/10.1109/ACCESS.2024.3351186

Hamzah, W. M. A. F. W., Yusoff, M. H., Ismail, I., & Ismail, N. (2021). Predicting Students' Behavioural Engagement in Microlearning using Learning Analytics Model. In *E-learning Methodologies: Fundamentals, technologies and applications* (pp. 53–78). Institution of Engineering and Technology. https://doi.org/10.1049/PBPC040E_ch3

Haque, R., Quek, A., Ting, C.-Y., Goh, H.-N., & Hasan, M. R. (2024). Classification Techniques Using Machine Learning for Graduate Student Employability Predictions. *International Journal on Advanced Science, Engineering and Information Technology*, *14*(1), 45–56. https://doi.org/10.18517/ijaseit.14.1.19549

Hasan, N., Adam, M. B., Mustapha, N., & Bakar, M. R. A. (2012). Sensitivity of Missing Values in Classification Tree for Large Sample. *AIP Conference Proceedings*, *1450*, 374–379. https://doi.org/10.1063/1.4724171

Hashim, R. A., Lim, H.-E., Jafar, M. F., Shanmugam, S. K. S., & Bukhari, N. (2024). Statistical Identification of Predictors of Dropout in Secondary Education: Evidence from Malaysia. *Journal of the Asia Pacific Economy*, *0*(0), 1–27. https://doi.org/10.1080/13547860.2024.2306673

Hassan, H., Ahmad, N. B., & Anuar, S. (2020). Improved Students' Performance Prediction for Multi-class Imbalanced Problems using Hybrid and Ensemble Approach in Educational Data Mining. *Journal of Physics: Conference Series*, *1529*(5). https://doi.org/10.1088/1742-6596/1529/5/052041

Hassan, H., Ahmad, N. B., & Sallehuddin, R. (2021). An Empirical Study to Improve Multiclass Classification Using Hybrid Ensemble Approach for Students' Performance Prediction. *Lecture Notes in Electrical Engineering*, *724*, 551–561. https://doi.org/10.1007/978-981-33-4069-5_45

Hassan, H., Anuar, S., & Ahmad, N. B. (2019). Students' Performance Prediction Model Using Meta-classifier Approach. In *Communications in Computer and Information Science* (Vol. 1000). Springer International Publishing. https://doi.org/10.1007/978-3-030-20257-6_19

Hoe, A. C. K., Ahmad, M. S., Hooi, T. C., Shanmugam, M., Gunasekaran, S. S., Cob, Z. C., & Ramasamy, A. (2013). Analyzing Students Records to Identify Patterns of Students' Performance. *2013 International Conference on Research and Innovation in*

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Information Systems (ICRIIS), 544–547. IEEE. https://doi.org/10.1109/ICRIIS.2013.6716767

Hong, C. M., Ch'ng, C. K., & Roslan, T. R. N. (2023). Predicting Students' Inclination to TVET Enrolment Using Various Classifiers. *Pertanika Journal of Science and Technology*, *31*(1), 475–493. https://doi.org/10.47836/pjst.31.1.28

Husin, W. Z. W., Zain, M. N. M., Zahan, N. A. N., Adam, P. N. A., & Aziz, N. A. (2022). Performance of Decision Tree and Neural Network Approach in Predicting Students' Performance. *International Journal of Academic Research in Business and Social Sciences*, *12*(6), 1252–1264. https://doi.org/10.6007/IJARBSS/v12-i6/13867

Isha, D. S. N. S. B. A., & Hashim, S. R. B. M. (2022). Application of Multiple Linear Regression in Identifying Contributing Factors in Students' Academic Achievement. In N. Wahi, M. A. Mohd Safari, R. Hasni, F. Abdul Razak, I. Gafurjan, & A. Fitrianto (Eds.), *Proceedings of the International Conference on Mathematical Sciences and Statistics 2022 (ICMSS 2022)* (pp. 364–377). Dordrecht: Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-014-5_32

Ismail, M. H., Razak, T. R., Noor, N. M., & Aziz, A. A. (2024). Evaluating Machine Learning Algorithms for Predicting Financial Aid Eligibility: A Comparative Study of Random Forest, Gradient Boosting and Neural Network. *Proceedings of the 2024 18th International Conference on Ubiquitous Information Management and Communication, IMCOM 2024*, 1–6. https://doi.org/10.1109/IMCOM60618.2024.10418450

Jamil, J. M., Pauzi, N. F. M., & Nee, I. N. M. S. (2018). An Analysis on Student Academic Performance by Using Decision Tree Models. *Journal of Social Sciences Research*, *2018*(Special Issue 6), 615–620. https://doi.org/10.32861/jssr.spi6.615.620

Khamis, S., Ahmad, M., Ahmad, A., & Ahmad, M. N. (2022). Internet Use Behaviour Model for Predicting Students' Performance. *Expert Systems*, *39*(8), 1–21. https://doi.org/10.1111/exsy.12999

Kumaran, S. R., Yusuf, L. M., Othman, M. S., & Yunianta, A. (2020). Educational Business Intelligence Framework Visualizing Significant Features using Metaheuristic Algorithm and Feature Selection. *2019 International Conference on Advances in the Emerging Computing Technologies, AECT 2019*. https://doi.org/10.1109/AECT47998.2020.9194221

Law, T.-J., Ting, C.-Y., Goh, H.-N., Ng, H., & Quek, A. (2024). Factors of Pre-University Study in Influencing Graduate on Time. *2024 16th International Conference on Advanced Computational Intelligence, ICACI 2024*, (Icaci), 36–42. https://doi.org/10.1109/ICACI60820.2024.10537019

Law, T.-J., Ting, C.-Y., Ng, H., Goh, H.-N., & Quek, A. (2024). Ensemble-SMOTE: Mitigating Class Imbalance in Graduate on Time Detection. *Journal of Informatics and Web Engineering*, *3*(2), 229–250. https://doi.org/10.33093/jiwe.2024.3.2.17

Lestari, W., Abdullah, A. S., Amin, A. M. A., Nurfaridah, Sukotjo, C., Ismail, A., … Utomo, C. P. (2024). Artificial Intelligence to Predict Pre-clinical Dental Student Academic Performance based on Pre-university Results: A Preliminary Study. *Journal of Dental Education*, (August). https://doi.org/10.1002/jdd.13673

Looi, Z. N., Song, P. C., Lim, H. T., & Looi, S. Y. (2024). A Case Study via Bayesian Network: Investigating Factors Influencing Student Academic Performance in Online Teaching and Learning During COVID-19 Pandemic. *Lecture Notes on Data Engineering and Communications Technologies*, *191*(June), 303–317. https://doi.org/10.1007/978-981-97-0293-0_23

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Lumius, L. D., & Asli, M. F. (2021). OBEInsights: Visual Analytics Design for Predictive OBE Knowledge Generation. *International Journal of Advanced Computer Science and Applications*, *12*(12), 895–901. https://doi.org/10.14569/IJACSA.2021.01212108

Lye, C.-T., Ng, L.-N., Hassan, M. D., Goh, W.-W., Law, C.-Y., & Ismail, N. (2010). Predicting Pre-university Student's Mathematics Achievement. *Procedia - Social and Behavioral Sciences*, *8*(5), 299–306. https://doi.org/10.1016/j.sbspro.2010.12.041

Mahmud, N., Pazil, N. S. M., & Azman, N. A. N. (2022). The Significant Factors Affecting Students' Academic Performance in Online Class: Multiple Linear Regression Approach. *Jurnal Intelek*, *17*(2), 1–11. https://doi.org/10.24191/ji.v17i2.17896

Makhtar, M., Nawang, H., & Shamsuddin, S. N. W. (2017). Analysis on Students Performance Using Naïve. *Journal of Theoretical and Applied Information Technology*, *31*(16), 3993–4000. Retrieved from www.jatit.org

Ministry of Education. (2013). Malaysia Education Blueprint 2013-2025 (Preschool to Post-Secondary Education). In *Kementerian Pendidikan Malaysia* (Vol. 1).

Ministry of Education. (2015). *Malaysia Education Blueprint 2015-2025 (Higher Education)* (Vol. 1). https://doi.org/10.1088/1751-8113/44/8/085201

Mohamad, N., Ahmad, N. B., & Jawawi, D. N. A. (2018). Malaysia MOOC: Improving Low Student Retention with Predictive Analytics. *International Journal of Engineering and Technology(UAE)*, *7*(2), 145–152. https://doi.org/10.14419/ijet.v7i2.29.13305

Nawai, S. N. M., Saharan, S., & Hamzah, N. A. (2021). An Analysis of Students' Performance Using CART Approach. *AIP Conference Proceedings*, *2355*, 060009. https://doi.org/10.1063/5.0053388

Nawang, H., Makhtar, M., & Hamzah, W. M. A. F. W. (2022). Comparative Analysis of Classification Algorithm Evaluations to Predict Secondary School Students' Achievement in Core and Elective Subjects. *International Journal of Advanced Technology and Engineering Exploration*, *9*(89), 430–445. https://doi.org/10.19101/IJATEE.2021.875311

Nawang, H., Makhtar, M., & Shamsuddin, S. N. W. (2018). Classification Model and Analysis on Students' Performance. *Journal of Fundamental and Applied Sciences*, *9*(6S), 869. https://doi.org/10.4314/jfas.v9i6s.65

Ng, K., Hoo, M.-H., Nair, M. B., & Khor, K.-C. (2023). Predicting Student Performance in Final Year Project Using Data Mining Classification Techniques. *AIP Conference Proceedings*, *2808*(May), 040008. https://doi.org/10.1063/5.0133061

Noor, M. M., Kadirgama, K., Rahman, M. M., Bakar, R. A., & Ibrahim, A. (2010). Analysis of Entry Qualification for Student Performance: An Intelligence Approach. *Journal of Engineering and Technology*, *1(1)*(December 2013), 53–62.

Norwawi, N. M., Hibadullah, F., Abdusalam, S. F., & Shuaibu, B. M. (2009). Classification of Students' Performance in Computer Programming Course According to Learning Style. *2009 2nd Conference on Data Mining and Optimization*, (December 2015), 37–41. IEEE. https://doi.org/10.1109/DMO.2009.5341912

Othman, W. N. A. W., & Abdullah, A. (2020). Student Learning Progress as Predictor for Graduate Employability Performance. *IOP Conference Series: Materials Science and Engineering*, *769*(1). https://doi.org/10.1088/1757-899X/769/1/012019

Othman, W. N. A. W., Abdullah, A., & Romli, A. (2020). Predicting Graduate Employability based on Program Learning Outcomes. *IOP Conference Series: Materials Science and Engineering*, *769*(1). https://doi.org/10.1088/1757-899X/769/1/012018

Othman, Z., Shan, S. W., Yusoff, I., & Kee, C. P. (2018). Classification Techniques for Predicting Graduate Employability. *International Journal on Advanced Science,*

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Engineering and Information Technology, *8*(4–2), 1712–1720. https://doi.org/10.18517/ijaseit.8.4-2.6832

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*(March). https://doi.org/10.1016/j.ijsu.2021.105906

Pauzi, W. N. D. W., Hasan, H., & Mahmud, Z. (2021). Supervised and Unsupervised Data Mining Techniques on Employability of Public Higher Learning Institute Graduates in Malaysia. *Journal of Physics: Conference Series*, *2084*(1). https://doi.org/10.1088/1742-6596/2084/1/012004

Rahim, A. A. A., & Buniyamin, N. (2022). Predicting Engineering Students' Academic Performance using Ensemble Classifiers- A Preliminary Finding. *Journal of Electrical & Electronic Systems Research*, *20*(APR2022), 92–101. https://doi.org/10.24191/jeesr.v20i1.013

Rahim, A. A. A., & Buniyamin, N. (2023). Mitigating Imbalanced Classification Problems in Academic Performance with Resampling Methods. *Journal of Electrical & Electronic Systems Research*, 45–56. https://doi.org/10.24191/jeesr.v23i1.006

Rahman, N. A. A., Tan, K. L., & Lim, C. K. (2017). Predictive Analysis and Data Mining among the Employment of Fresh Graduate Students in HEI. *AIP Conference Proceedings*, *1891*. https://doi.org/10.1063/1.5005340

Rahman, N. A. B. A., Tan, K. L., & Lim, C. K. (2017). Supervised and Unsupervised Learning in Data Mining for Employment Prediction of Fresh Graduate Students. *Journal of Telecommunication, Electronic and Computer Engineering*, *9*(2), 155–161.

Rahman, N. H. A., Sulaiman, S. A., & Ramli, N. A. (2023). The Development of a Predictive Model for Students' Final Grades Using Machine Learning Techniques. *Data Analytics and Applied Mathematics (DAAM)*, *4*(1), 40–48. https://doi.org/https://doi.org/10.15282/daam.v4i1.9591

Rahman, N. H. A., Sulaiman, S. A., & Ramli, N. A. (2024). The Predictive Modelling of Student Academic Performance Using Machine Learning Approaches. In *International Conference on Soft Computing and Data Mining* (pp. 379–389). https://doi.org/10.1007/978-3-031-66965-1_37

Raihana, Z., & Nabilah, A. M. F. (2018). Classification of Students Based on Quality of Life and Academic Performance By Using Support Vector Machine. *Journal of Academia UiTM Negeri Sembilan*, *6*(1), 45–52.

Rani, M. H., & Embong, A. (2013). Predicting Student Performance in Object Oriented Programming Using Decision Tree : A Case at Kolej Poly-Tech Mara, Kuantan. *3rd International Conference on Software Engineering & Computer Systems (ICSECS - 2013)*.

Razak, R. A., Omar, M., & Ahmad, M. (2018). A Student Performance Prediction Model Using Data Mining Technique. *International Journal of Engineering and Technology(UAE)*, *7*(2), 61–63. https://doi.org/10.14419/ijet.v7i2.15.11214

Razali, M. N., Zakariah, H., Hanapi, R., & Rahim, E. A. (2022). Predictive Model of Undergraduate Student Grading Using Machine Learning for Learning Analytics. *Proceedings - 2022 4th International Conference on Computer Science and Technologies in Education, CSTE 2022*, 260–264. https://doi.org/10.1109/CSTE55932.2022.00055

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013
Roslan, M. H. Bin, & Chen, C. J. (2023). Predicting Students' Performance in English and Mathematics using Data Mining Techniques. *Education and Information Technologies*, *28*(2), 1427–1453. https://doi.org/10.1007/s10639-022-11259-2

Sabri, M. Z. M., Majid, N. A. A., Hanawi, S. A., Talib, N. I. M., & Yatim, A. I. A. (2023). Prediction Model based on Continuous Data for Student Performance using Principal Component Analysis and Support Vector Machine. *TEM Journal*, *12*(2), 1201–1210. https://doi.org/10.18421/TEM122-66

Sabri, N. M., & Hamrizan, S. F. A. (2023). Prediction of MUET Results Based on K-Nearest Neighbour Algorithm. *Annals of Emerging Technologies in Computing*, *7*(5), 50–59. https://doi.org/10.33166/AETiC.2023.05.005

Sahrin, S., & Muhammad, N. (2024). Identifying Factors that Influence Students' Performance through Multiple Linear Model. *AIP Conference Proceedings*, *2895*(1), 090003. https://doi.org/10.1063/5.0193364

Samsudin, N. A. M., Shaharudin, S. M., Sulaiman, N. A. F., Fuad, M. F. M., Zulfikri, M. F., & Zainuddin, N. H. (2021). Modeling Student's Academic Performance During Covid-19 Based on Classification in Support Vector Machine. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(5), 1798–1804. https://doi.org/10.17762/turcomat.v12i5.2190

Samsudin, N. A. M., Shaharudin, S. M., Sulaiman, N. A. F., Ismail, S., Mohamed, N. S., & Husin, N. H. M. (2022). Prediction of Student's Academic Performance during Online Learning Based on Regression in Support Vector Machine. *International Journal of Information and Education Technology*, *12*(12), 1431–1435. https://doi.org/10.18178/ijiet.2022.12.12.1768

Sangodiah, A., Beleya, P., Muniandy, M., Heng, L. E., & SPR, C. R. (2015). Minimizing Student Attrition in Higher Learning Institutions in Malaysia using Support Vector Machine. *Journal of Theoretical and Applied Information Technology*, *71*(3), 377–385.

Sani, N. S., Nafuri, A. F. M., Othman, Z. A., Nazri, M. Z. A., & Mohamad, K. N. (2020). Drop-Out Prediction in Higher Education Among B40 Students. *International Journal of Advanced Computer Science and Applications*, *11*(11), 550–559. https://doi.org/10.14569/IJACSA.2020.0111169

Sapaat, M. A., Mustapha, A., Ahmad, J., Chamili, K., & Muhamad, R. (2011). A Classification-Based Graduates Employability Model for Tracer Study by MOHE. *Communications in Computer and Information Science*, *188 CCIS*(PART 1), 277–287. https://doi.org/10.1007/978-3-642-22389-1_25

Sembiring, S., Zarlis, M., Hartama, D., S, R., & Wani, E. (2011). Prediction of Student Academic Performance By an Application of Data Mining Techniques. *2011 International Conference on Management and Artificial Intelligence*, *6*(January), 110–114. IACSIT Press.

Shariff, S. S. R., Rodzi, N. A. M., Rahman, K. A., Zahari, S. M., & Deni, S. M. (2016). Predicting the "Graduate on Time (GOT)" of PhD Students Using Binary Logistics Regression Model. *AIP Conference Proceedings*, *1782*(February 2019), 050015. https://doi.org/10.1063/1.4966105

Shukor, N. A., Tasir, Z., & Meijden, H. Van der. (2015). An Examination of Online Learning Effectiveness Using Data Mining. *Procedia - Social and Behavioral Sciences*, *172*, 555–562. https://doi.org/10.1016/j.sbspro.2015.01.402

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Siraj, F. (2016). Modeling Academic Achievement of UUM Graduate Using Descriptive and Predictive Data Mining. In *Lecture Notes in Electrical Engineering* (Vol. 362, pp. 609–620). https://doi.org/10.1007/978-3-319-24584-3_52

Siraj, F., & Bakar, N. A. A. (2019). Identifying Patterns of Students Academic Performance from Tracer Evaluation using Descriptive Data Mining. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 2), 187–191. https://doi.org/10.35940/ijrte.B1034.0782S219

Siraj, F., Yusoff, N., & Ali, N. M. (2006). Exploring Hidden Relationships within Students ' Data Using Neural Network and Logistic Regression. *Proceedings of Knowledge Management International Conference & Exhibition (KMICE)*, 162–168. School of Computing, College of Arts and Sciences, Universiti Utara Malaysia.

Suhaimi, N. M., Abdul-Rahman, S., Mutalib, S., Hamid, N. H. A., & Malik, A. M. A. (2019). Predictive Model of Graduate-On-Time Using Machine Learning Algorithms. In *Communications in Computer and Information Science* (Vol. 1100). Springer Singapore. https://doi.org/10.1007/978-981-15-0399-3_11

Suliman, N. A., Abidin, B., Manan, N. A., & Razali, A. M. (2014). Predicting Students' Success at Pre-University Studies Using Linear and Logistic Regressions. *AIP Conference Proceedings*, *1613*(Soric 2013), 306–316. https://doi.org/10.1063/1.4894355

Tan, C. J., Lim, T. Y., Liew, T. K., & Lim, C. P. (2022). An Intelligent Tool for Early Drop-out Prediction of Distance Learning Students. *Soft Computing*, *26*(12), 5901–5917. https://doi.org/10.1007/s00500-021-06604-5

Teoh, C.-W., Ho, S.-B., Dollmat, K. S., & Tan, C.-H. (2022). Ensemble-Learning Techniques for Predicting Student Performance on Video-Based Learning. *International Journal of Information and Education Technology*, *12*(8), 741–745. https://doi.org/10.18178/ijiet.2022.12.8.1679

Teoh, C.-W., Ho, S.-B., Dollmat, K. S., & Tan, C.-H. (2023). Machine Learning Prediction Model for Early Student Academic Performance Evaluation in Video-Based Learning. *International Journal of Membrane Science and Technology*, *10*(2), 1529–1544. https://doi.org/10.15379/ijmst.v10i2.1822

Ting, C.-Y., Cheah, W.-N., & Ho, C. C. (2013). Student Engagement Modeling using Bayesian Networks. *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, 2939–2944. https://doi.org/10.1109/SMC.2013.501

Wan Yaacob, W. F., Mohd Sobri, N., Nasir, S. A. M., Wan Yaacob, W. F., Norshahidi, N. D., & Wan Husin, W. Z. (2020). Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques. *Journal of Physics: Conference Series*, *1496*(1). https://doi.org/10.1088/1742-6596/1496/1/012005

Wook, M., Yahaya, Y. H., Wahab, N., Isa, M. R. M., Awang, N. F., & Seong, H. Y. (2009). Predicting NDUM Student's Academic Performance Using Data Mining Techniques. *2009 Second International Conference on Computer and Electrical Engineering*, *2*, 357–361. IEEE. https://doi.org/10.1109/ICCEE.2009.168

Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised Data Mining Approach for Predicting Student Performance. *Indonesian Journal of Electrical Engineering and Computer Science*, *16*(3), 1584–1592. https://doi.org/10.11591/ijeecs.v16.i3.pp1584-1592

Yusof, M. H. M., & Khalid, I. A. (2021). Precision Education Reviews: A Case Study on Predicting Student's Performance using Feed Forward Neural Network. *2021*

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 41 (December 2025) PP. 204-224
DOI: 10.35631/JISTM.1041013

International Conference of Technology, Science and Administration, ICTSA 2021, (Ld), 3–6. https://doi.org/10.1109/ICTSA52017.2021.9406525

Yusof, R., Hashim, N., Rahman, N. A., Yunus, S. Y. M., & Fadzillah, N. A. A. (2022). Academic Performance Prediction Model Using Classification Algorithms: Exploring the Potential Factors. *International Journal of Academic Research in Progressive Education and Development*, *11*(3), 706–724. https://doi.org/10.6007/IJARPED/v11-i3/14753

Zahruddin, N. A. binti M., Kamarudin, N. D., Jusoh, R. M., Fataf, N. A. A., & Hidayat, R. (2023). Case Study: Using Data Mining to Predict Student Performance Based on Demographic Attributes. *JOIV : International Journal on Informatics Visualization*, *7*(4), 2460. https://doi.org/10.30630/joiv.7.4.02454

Zakaria, F., Kar, S., Abdullah, R., & Ismail, S. I. (2019). ANN Application: To Study Effect of Literacy and Numeracy Skills on Electrical Engineering Students' Performance. *Advances in Computing and Intelligent System (ACIS)*, *1*(2), 1–5.

Zakaria, N. A., Ahmad, T., Awang, S. R., & Safar, A. (2021). Determination of Huffaz Academic Achievement Using Binary Logistic Regression Model. *Journal of Physics: Conference Series*, *1988*(1). https://doi.org/10.1088/1742-6596/1988/1/012104

Zaki, S. M., Razali, S., Kader, M. A. R. A., Laton, M. Z., Ishak, M., & Burhan, N. M. (2024). Predicting Students' Performance at Higher Education Institutions Using a Machine Learning Approach. In *Kybernetes*. https://doi.org/10.1108/K-12-2023-2742

Zulfikri, M. F., Shaharudin, S. M., Rajak, N. A. A., & Ibrahim, M. S. (2021). Predictive Analytics on Academic Performance in Higher Education Institution during COVID-19 using Regression Model. *International Journal of Biology and Biomedical Engineering*, *15*, 184–189. https://doi.org/10.46300/91011.2021.15.21