Journal of Quality Measurement and Analysis *JQMA* 21(4) 2025, 389-409
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2104.2025.20

LEARNING ANALYTICS OF ONLINE STUDENTS PERFORMANCE IN MATHEMATICS USING BAYESIAN NETWORK

(*Analitik Pembelajaran Prestasi Matematik Pelajar dalam Talian Menggunakan Rangkaian Bayesian*)

NURULHUDA RAMLI* & MOHD TAHIR ISMAIL

ABSTRACT

In recent years, there has been growing interest in online education, including e-learning, Massive Open Online Courses (MOOCs), Intelligent Tutoring System (ITS) and university-level distance learning. These online environments generate vast amounts of data, such as activity logs, course interaction data and assessment results, making learning analytics essential to predicting student performance. This study examines the efficacy of the probability-based model, Bayesian Networks (BNs), in predicting academic performance using learning analytics data collected through the Learning Management Systems (LMS). It focuses on how BN is capable of predicting student performance in an online learning environment by modeling complex relationships among various learning analytics factors that contribute to academic success. Using LMS data from Universiti Sains Malaysia's distance learning Mathematics course, the study incorporates key learning analytics variables such as engagement metrics, resource utilization, self-directed learning activities and assessment, and academic performance to develop a BN-based predictive model. BN model revealed that low engagement significantly hinders academic success, demonstrating its potential for early intervention and educational improvement. The model performance was measured using classification metrics such as accuracy, precision, recall, and F1-score. The developed model shows overall good performance, marked by strong precision and balanced recall in predicting the target classes with some variability. The results revealed that BN effectively captured dependencies among key learning analytics variables, providing actionable insights for designing personalized interventions in online education.

Keywords: Bayesian network; learning analytics; student performance; online learning; predictive modelling; Mathematics education

ABSTRAK

Kebelakangan ini menyaksikan peningkatan minat dalam pendidikan dalam talian, termasuk e-pembelajaran, Kursus Dalam Talian Terbuka Besar-besaran (MOOCs), Sistem Tutor Pintar (ITS), dan pembelajaran jarak jauh peringkat universiti. Persekitaran dalam talian ini menghasilkan sejumlah besar data, seperti log aktiviti, data interaksi kursus, dan hasil penilaian menjadikan analitik pembelajaran penting untuk meramalkan prestasi pelajar. Kajian ini mengkaji keberkesanan model berasaskan kebarangkalian, Rangkaian Bayesian (BN), dalam meramalkan prestasi akademik menggunakan data analitik pembelajaran yang dikumpulkan melalui Sistem Pengurusan Pembelajaran (LMS). Ia memberi tumpuan kepada bagaimana BN mampu meramalkan prestasi pelajar dalam persekitaran pembelajaran dalam talian dengan memodelkan hubungan kompleks antara pelbagai faktor analitik pembelajaran yang menyumbang kepada kejayaan akademik. Menggunakan data LMS kursus Matematik pembelajaran jarak jauh Universiti Sains Malaysia, kajian ini menggabungkan pemboleh ubah analitik pembelajaran seperti metrik penglibatan, penggunaan sumber, aktiviti pembelajaran kendiri dan penilaian, serta prestasi akademik untuk membangunkan model ramalan berasaskan BN. Model BN menunjukkan bahawa kurangnya penglibatan secara signifikan menghalang kejayaan akademik, menggambarkan potensi untuk intervensi awal dan peningkatan pendidikan. Prestasi model diukur menggunakan metrik klasifikasi seperti ketepatan, kepersisan, ingatan semula, dan skor-F1. Model yang dibangunkan ini memaparkan prestasi keseluruhan yang baik dalam meramalkan kelas sasaran, melalui nilai kepersisan yang tinggi dan ingatan semula yang seimbang, di samping terdapat sedikit kebolehubahan. Dapatan kajian menunjukkan bahawa BN berjaya menggambarkan kebergantungan antara pemboleh ubah utama analitik pembelajaran, sekaligus memberikan pemahaman yang boleh diambil untuk merekabentuk intervensi peribadi dalam pendidikan dalam talian.

Kata kunci: rangkaian Bayesian; analitik pembelajaran; prestasi pelajar; pembelajaran dalam talian; pemodelan ramalan, pendidikan Matematik

1. Introduction

E-learning systems in higher education have transformed conventional pedagogical approaches by using advanced online technology to provide remote access to educational content and foster interactive and personalized learning experiences. Some of the key components of these systems include Learning Management Systems (LMS) such as Modular Object-Oriented Dynamic Learning Environment (Moodle), Virtual Learning Environment (VLE), Massive Open Online Courses (MOOCs) and adaptive learning technologies that customize learning instruction to meet individual student needs. These systems bring numerous benefits such as enhanced accessibility and convenience (Pihlajamaa *et al.* 2016; Alshahrani 2023), cost-effectiveness (Pakdaman *et al.* 2019), personalized learning trajectories (Jose *et al*. 2024), improved communication (Pihlajamaa *et al.* 2016), and increased engagement through use of multimedia resources and gamification (Kato Nabirye 2025). The effectiveness of e-learning content delivery is enriched by its use of a wide array of media, including text, audio, images, animation and streaming videos. With rapid technology and the advancement of learning systems, devices like smartphones and tablets have gained significant importance in classrooms, facilitating mobile learning (M-learning) that offers learning flexibility on the go (Lazaro & Duart 2023).

The rise of e-learning systems in higher education has paved the way for the contribution of learning analytics, which play a crucial role in transforming every aspect of the learning process. According to Siemens and Long (2011), learning analytics involves the systematic collection, analysis, and interpretation of student data from these digital learning environments to identify patterns, predict outcomes, and enhance personalized learning experiences. This process includes pulling a wide range of data sources, from LMS to online assessments to logs of student interaction — anything and everything that provides a detailed and comprehensive view of how students learn. Learning analytics assists educators and education providers by offering actionable insights into course effectiveness, which helps continuous improvements in curriculum design and teaching strategies (Wong *et al.* 2025). For instance, by analyzing student engagement metric like how long student spend with learning materials or their participation in online discussions, educators can identify where students might be having difficulties and modify their teaching approaches accordingly (Baker & Inventado 2014). Furthermore, learning analytics also plays a vital role in informing institutional capacity in decision-making by providing data-driven insights into student success and resource allocation (Daniel 2015). By analyzing this data, institutions can uncover important trends and patterns that guide strategic planning and shape educational policies.

Previous research focused on modeling, measuring and predicting student performance using learning analytics have explored a variety of predictive models, particularly machine learning (ML) techniques used to enhance educational results. Studies have shown that predictive models can estimate students' success using state-based and event-driven data, such as demographics, past performance and engagement with online learning environments (Almalawi *et al*. 2024). These models can also identify students who are at risk of facing challenges early on, allowing for timely interventions and personalized support (Romero & Ventura 2007). Despite these advances, significant challenges remain particularly in achieving truly personalized evaluation. The realization of personalization is hindered by data quality, algorithmic bias, and the need for models that can adapt to diverse and evolving learner profiles (Almalawi *et al*. 2024). Traditional predictive methods lack transparency, making it difficult for educators to understand or trust predictions (Bird 2021). To address these limitations, leveraging Artificial Intelligence (AI) and learning analytics can improve prediction accuracy by incorporating dynamic learning behavior and real-time assessments, allowing institutions to monitor student progress, identify learners at risk and provide timely interventions to improve academic success (Siemens & Baker 2012; Ouyang *et al*. 2023).

One such AI-probabilitistic approach, Bayesian Networks (BNs), has emerged as a promising approach in this field for their ability to handle uncertainties in student learning trajectories and improve the adaptability of personalized feedback mechanisms. BNs work by representing knowledge as a network of probabilistic relationships, offering a strong framework to model the complex interactions between various educational factors, such as student performance, learning resources, teaching strategies and academic results (Chen *et al.* 2024; Looi *et al.* 2023). Unlike traditional predictive models which often struggle with uncertainty and interconnected nature of variables, BNs manage these complexities effectively, making them particularly well-suited for educational assessment (Conati *et al.* 2002; Culbertson 2016). This is particularly important in education, where elements like student motivation, prior knowledge, and the learning environment interact in complex and sometimes unpredictable ways. This strength makes BNs highly valuable for evaluating the impact of adaptive learning interventions. Research has shown that BNs can accurately predict student progress, model learning processes to estimate different learning states, and ultimately provide targeted support and early intervention for at-risk students (Kondo & Hatanaka 2018; 2019).

Another major strength of BNs is their ability to integrate prior expert knowledge with observed data. This makes them highly effective for a range of learning tasks such as predicting student performance, grouping similar learning behaviors, and spotting unusual patterns (Heckerman 1997; Wang & Han 2016; Kitson *et al.* 2023). By combining domain expertise with real-world evidence, these models achieve greater accuracy and are easier to interpret. This feature is especially important in education, where expert insights help ensure that the relationships modeled truly reflect real-world causes and support meaningful decision-making for personalized learning and early intervention systems (Looi *et al.* 2023; Chen *et al.* 2024).

In addition, a key advantage of BNs lies in their inference process, which significantly support causal inference and prediction by enabling the learning and representation of complex cause-and-effect relationships in data. This inference mechanism allows beliefs to be dynamically updated as new evidence comes in, making BNs a perfect fit for adaptive learning environments that rely on continuous feedback and personalization (Wang & Han 2016). This capability allows educators and researchers to move beyond simple correlations to rigorously investigate how specific educational practices directly impact student achievement (Chen *et al.* 2024). This also includes simulating how different teaching strategies might play out, helping to predict possible outcomes, supporting better decision-making and tailored intervention planning (Delen *et al*. 2020; Jiang et *al.* 2023).

BNs have been increasingly applied in STEM education at the higher education level, especially for mathematics online learning, to model student learning, personalize teaching and improve educational outcomes. Foundational studies such as those by Millán *et al*. (2010; 2013), Seffrin *et al*. (2016), Käser *et al.* (2017) and Fan *et al.* (2021) demonstrate how BN models can assess student knowledge, identify misconceptions, support adaptive learning and course recommendation systems in online environments. However, most of this research focus on translating mathematics knowledge into learning pathways that cater to individual needs but do not fully incorporate learning analytics data, which is critical for addressing the diverse and dynamic needs of learners. Furthermore, much of the existing BN research concentrates on single data modalities such as log files, clickstream data, or grades, with limited exploration of combining diverse data sources, including video, textual or sensor information. These multimodal data sources are crutial for capturing the full complexity of learner behaviors (Liu *et al.* 2022; Zheng 2025). There is also a notable gap in domain-specific BN models tailored specifically for online mathematics education, which poses unique challenges such as developing problem-solving skills and assessing mastery in a step-by-step manner.

Therefore, the objective of this paper is to develop a BN model that leverages learning analytics data to achieve a more comprehensive understanding and prediction of student performance, particularly within the context of online mathematics in higher education. This model aims to effectively capture uncertainties and causal relationships in student learning behaviors, thereby providing more accurate assessments of factors influencing academic success and supporting personalized interventions.

2. Related Work

This section explores existing research on predicting student performance in online learning using learning analytics, focusing on the context of data-driven education and predictive modeling. The foundations for understanding BNs will then be introduced, as they serve as the core methodological approach for the research presented in this paper. Finally, the application of BNs in online mathematics within higher education is presented.

2.1. *Learning analytics*

Learning analytics has emerged as a vital area in online education, focused on extracting meaningful insights about teaching and learning from data. The origins of learning analytics can be traced back to the early 2000s, with its foundation built upon educational data mining (EDM) and academic analytics (Siemens 2013). While it initially concentrated on helping institutions make decision, the rapid growth of digital learning platforms has expanded boundaries of learning analytics far beyond that, reaching areas like student engagement, personalized learning, and predicting student performance. At its core, learning analytics involves collecting, analyzing and interpreting data related to students, all with the goal of maximizing learning outcomes and improving educational practices. More recent research points to its growing role in adaptive learning systems, wherein AI-based models personalize instructional content based on individual student's performance (Ouyang *et al*. 2023).

Learning analytics leverages various key variables to assess, predict and improve student performance in higher education. These include academic achievements, behavioral data, levels of engagement, demographic and psychographic characteristics, institutional influences and preferred learning methods. Some specific types of data involved are (Baker & Inventado 2014):

- Academic test scores
- Class grades/level
- Demographic and psychographic data
- Data of learning styles, characteristics or preferences
- LMS/ Content Management System (CMS) activity data
- Survey data

Academic performance such as grades, exam scores, coursework completion and assessment-related data remains as crutial factor in learning analytics. Research has shown that predictive models based on these academic measures are effective at identifying students who may be at risk, enabling early interventions that can improve outcomes. For example, Qiu *et al.* (2024) developed a dual-mode grade prediction architecture that enables grade prediction solely from past semester grades to identify at-risk students. The architecture uses a dual-mode approach that first incorporates a weighted loss function into an LSTM model, followed by a short-term Gated LSTM to facilitate early intervention strategies. Similarly, Lyn *et al*. (2024) developed a unified Learning Analytics Framework that uses clustering algorithms combined with network analysis to provide actionable insights. This framework aims to offer a comprehensive understanding of students' academic performance and highlight key learning pathways.

LMS log data plays a vital role in the early prediction of academic outcomes by tracking important student behaviors. Video-viewing behaviors—such as time spent watching instructional videos, repeated viewing, and active interactions like pausing or seeking—strongly correlate with student comprehension and engagement (Brinton *et al*. 2016; Lu *et al*. 2018). Likewise, early quiz attempts and scores give immediate feedback on how well students understand the course material, making it easier to identify those who might be struggling and to predict final grades sooner (Lu *et al.* 2018; Zhao *et al.* 2023). Homework or assignment submissions and their timeliness further serve as strong behavioral indicators linked to academic success (Moreno-Marcos *et al.* 2020).

These predictive insights are deeply connected to broader patterns of student engagement within digital learning platforms. Engagement metrics—such as clicks on course materials, frequency of access, time spent on digital coursework, forum participation and navigation within the LMS—capture students' overall interaction with the learning environment (Lu *et al*. 2018). For instance, Wang & Yu (2025) utilized a machine learning approach for student performance prediction in online learning, focusing on various online behavioral indicators. These indicators included course registration and login frequency, resource monitoring time, resource utilization efficiency, and engagement metrics such as repeated learning resource access and forum browsing/replying. A study by Bellarhmouch *et al.* (2025) proposed an intelligent predictive model designed to classify students based on their participation and engagement in the VLE. This algorithm aims to predict whether a learner is engaged or not, utilizing input data that includes student information, exam scores, quiz results, and lesson activity. Furthermore, more advanced methods like deep learning have been used to analyze temporal LMS data, demonstrating high accuracy in predicting course performance based on engagement behavior (Chen & Cui 2020).

These studies underscore the importance of leveraging diverse data sources to gain a comprehensive understanding of student learning and inform effective educational interventions. By integrating academic performance data with LMS activity and student behavioral insights, educators can create a more comprehensive and predictive learning analytics model that enables more targeted, timely, and effective educational interventions.

2.2. *Predictive modeling in learning analytics*

Predictive modeling plays a crucial role in translating learning analytics data into actionable insights. Acording to Ferguson (2012), the learning analytics process encompasses data analysis, prediction, and subsequent adaptation or intervention. Within this process, predictive modeling has become a key focus, helping educators and institutions better understand learner behavior and improve outcomes. These models are generally built using statistical or ML techniques to detect patterns in educational data. Many institutions have employed predictive modeling as part of their learning analytics efforts to boost student success (Arnold & Pistilli 2012; Jayaprakash *et al*. 2014; Sulak & Koklu 2024).

The application of ML and data mining techniques have gained considerable attention for their ability to predict learning outcomes and identify at-risk students. Recent advancements have enabled the integration of diverse learning analytic variables into highly accurate predictive models. Numerous studies and systematic reviews highlight the use of various ML algorithms such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Tree (DT), Random Forests (RF) and ensemble methods (Namoun & Alshanqiti 2021; Luan & Tsai 2021; Rizwan *et al.* 2025).

Deep learning approaches have also shown strong potential. For example, Kim *et al.* (2018) introduced the GritNet model, which uses sequential event data from MOOCs to make early predictions about whether students will complete a course. It performed better than logistic regression, especially in the early weeks of a course. Similarly, Lee *et al.* (2021) developed two deep neural networks—one that used video-viewing behaviors to predict learning outcomes, and another based on exercise data to predict question accuracy. Both models effectively measured student performance using learning behaviors and responses.

In the context of hybrid learning during the COVID-19 pandemic, Wan *et al.* (2023) developed a powerful deep learning model that combined Bidirectional Long Short-Term Memory (LSTM), Global Average Pooling and TIME MASK components. This model excelled at making early predictions in hybrid learning environments. Another study by Li *et al.* (2024) created a system using a lightweight Gradient Boosting Machine (LightGBM) optimized with a Genetic Algorithm (GA), which led to measurable improvements in students' online learning performance, including grades, midterms and GPA. Recently, Wang and Yu (2025) used logistic regression based on eleven behavioral indicators from online learning, showing strong links between student engagement such as time spent and initiative, and academic outcomes.

Several studies have also compared different ML models across educational settings to determine which ones work best for predicting student performance. For example, Nespereira *et al.* (2016) analyzed two years of Moodle access logs in a blended learning environment and found that RF outperformed SVM in predicting student success, emphasizing the importance of time-based interaction data. In a separate study, Ahmed (2024) tested several classifiers including SVM, DT, Naive Bayes, and K-Nearest Neighbors (KNN) on e-learning data. After tuning, SVM achieved the highest prediction accuracy at 96%,-while Naive Bayes lagged due to its assumption of independent features. Similarly, Habti *et al.* (2025) evaluated RF, Logistic Regression, SVM and Linear Discriminant Analysis (LDA) using open university data and found that RF was the most accurate, achieving 91% accuracy.

Despite their success, many existing ML models still face challenges, particularly in terms of explain ability and the lack of probabilistic reasoning. This limits their ability to provide deep, evidence-based insights that are essential for designing effective, personalized learning interventions. In contrast, BNs offer a more comprehensive and holistic modeling approach by capturing complex causal relationships and accounting for uncertainty in student learning processes. This allows for more precise, actionable insights that support personalized interventions, making BNs especially well-suited for educational settings. The following section discusses the BN framework in detail.

2.3. *Bayesian Network*

Bayesian Network (BN) fall into the realm of several fields such as graph theory and probability (Koller & Friedman 2009). BN is utilized as a probabilistic modeling approach capable of providing a flexible representation of relationships among variables while maintaining a high level of interpretability (Pearl 1988). BN represents probability distribution by approximating a joint probabilistic structure through a Directed Acyclic Graph (DAG), i.e., a directed graph with-out cycles. The variables (or nodes in the network) can be in two categories, the parent nodes—variables that are causes of a particular node, and child node—the consequence of that node. The nodes with no parent nodes are called root nodes. The variables associated with the nodes can be discrete—take on a finite number of states, or continuous—an infinite range of possible values. The directional arches from parent to child nodes represent causal influences, quantified by the conditional probability values, which are represented in a Conditional Probability Table (CPT) for discrete variables. Each probability in a CPT represents the probability of a certain state in a child node given a single state or a set of states in parent nodes. For nodes that have no parents, the CPT will contain the prior probability values for discrete variables or marginal distribution for continuous variables.

Assume that a set of random variables $V = \{X_1, X_2, \dots, X_n\}$ and a DAG U=(V,D) where V is the set of variables in the network and D represents the directed edges. Each variable X_i is associated with a conditional probability distribution $P(X_i|Pa(X_i))$. The joint probability of all variables is given by:

$$\mathrm{P}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{P}(\mathrm{X}_\mathrm{i}|\mathrm{Pa}(\mathrm{X}_\mathrm{i})) \tag{1}$$

where $\mathrm{Pa}(\mathrm{X_i})$ represents the parent nodes of $\mathrm{X_i}$. The marginal probability of variable $\mathrm{Y_j}$ is obtained by summing over all possible values of other variables:

$$\mathrm{P}(\mathrm{Y_j}) = \sum_\mathrm{X} \mathrm{P}(\mathrm{X} = \mathrm{x_i})\mathrm{P}(\mathrm{Y} = \mathrm{y_j}|\mathrm{X} = \mathrm{x_i}). \tag{2}$$

The basic principle governing BN is Bayes' theorem given by;

$$\mathrm{P}(\mathrm{X_i}|\mathrm{Y_j}) = \frac{\mathrm{P}(\mathrm{X_i})\mathrm{P}(\mathrm{Y_j}|\mathrm{X_i})}{\mathrm{P}(\mathrm{Y})} \tag{3}$$

where $P(X_i|Y_y)$ is the posterior probability, $P(X_i)$ is the prior probability and $P(Y_j|X_i)$ is the likelihood.

There are three main stages to design and build the BN: variables structuring, quantification and inference. The first stage is to identify variables and their dependencies or cause-effect relationship. Structure learning is the process of defining the optimal configuration of a BN to capture relationships between variables effectively. There are two main approaches to constructing a BN structure: expert-driven and algorithmic. In the expert-driven approach, domain specialists manually define the network based on their knowledge of causal dependencies within the system. This method is particularly useful when data is limited or when prior knowledge is essential in guiding the model's structure (Russell & Norvig 2021). The second approach relies on algorithmic methods, where ML techniques infer the network structure from data (Murphy 2002).

Once the structure of a BN is determined, the next step is parameter estimation, which involves defining the prior probability (for each state in the parent node) and CPT values (for each state in the child node) for each node in the network. Two widely used approaches for parameter estimation are based on exact data, utilizing methods such as Maximum Likelihood Estimation (MLE) and Bayesian Estimation, or on expert judgment when data is unavailable, through a procedure known as expert elicitation (Jensen & Nielsen 2007).

The third stage is to make an inference on the variables when new information (called evidence) of any variables is entered in the network. When evidence is observed in one of the variables, it can propagate over the network and the probabilities of all other variables that are affected are updated as well. BNs involve two types of inference. The first is causal (or top-down) inference, in which probabilities in child nodes are computed based on evidence from parent nodes. The second is diagnostic (or bottom-up) inference, where evidence is observed in child nodes, and the goal is to infer the most likely causes in the parent nodes.

Various software packages are available for building BN, including Netica, Hugin, Analytica, GeNIe, and Bayes Net Toolbox. These tools facilitate the efficient construction, analysis, and visualization of BN, making the modeling process more accessible. In this study, GeNIe Modeler (BayesFusion 2023) is used to develop the BN model due to its high quality, flexible data-generation capabilities, user-friendly graphical interface, and free availability for academic users. It was developed by the Decision Systems Laboratory at the University of Pittsburgh and is part of the SMILE (Structural Modeling, Inference, and Learning Engine) library which allows integration with other applications.

2.4. *Bayesian Networks in online mathematics higher education*

BNs have increasingly been applied in online mathematics education to model student learning to better understand student learning, deliver personalized instruction, and improve outcomes in higher education. One of the early foundational studies by Millán *et al.* (2010) laid important groundwork in building Bayesian Student Model (BSM) for procedural mathematics domains, specifically modeling the Simplex algorithm used in linear programming. In their work, student knowledge was represented as a sequence of interconnected skills necessary for completing each step of the algorithm. This approach demonstrated how adaptive tutoring systems can tailor instruction to a student's mastery of complex, step-by-step mathematical procedures in an online learning setting.

Building on that foundation, Millán *et al.* (2013) integrated a Generic Bayesian Student Model (GBSM) into the Mathematics Education Project’s computerized testing system to diagnose student knowledge of first-degree algebraic equations. Their model was validated with 152 students who completed both computerized and written exams. Results showed strong alignment between BSM diagnoses and expert grading in written tests, but weaker agreement for computerized assessments. This study revealed challenges in applying BN diagnostics to automated online testing environments.

Seffrin *et al.* (2016) advanced BN applications by developing a Dynamic Bayesian Network (DBN) for step-based Intelligent Tutoring Systems (ITS) in assessing algebraic knowledge. Rather than only analyzing final answers, their model assessed each algebraic operation performed during problem-solving to infer understanding of both concepts and skills. Their model also relied on explicitly defined prerequisite relationships among knowledge components to support fine-grained learning assessments. This allowed for more accurate detection of misconceptions and enabled highly personalized feedback.

Further expanding BN capabilities, Käser *et al.* (2017) used DBNs to overcome the limitations of traditional Bayesian Knowledge Tracing (BKT). Their model simultaneously tracked multiple interconnected skills across various domains, including mathematics. Evaluated on large-scale datasets, the DBN approach proved more accurate at predicting students' knowledge states and helped design personalized instructional strategies in intelligent online tutoring systems.

Fan *et al*. (2021) tackled the challenge of course personalization in online math platforms by developing a Bayesian-based recommendation model. This model encoded causal relations between knowledge points and used both prior and posterior information to recommend optimal course selections. It outperformed traditional collaborative filtering methods, allowing learners to independently choose courses aligned with their current progress and personal preferences, ultimately supporting greater autonomy in online learning.

In Malaysia, research applying predictive modeling, particularly BNs to understand mathematics learning in higher education is still limited but gradually growing. To date, only one prominent study by Ong and Lim (2014) has applied BNs in the context of mathematics Malaysian higher education. Their research explored the factors influencing problem-solving among pre-university students and found that a poor understanding of mathematical symbols significantly reduced students' confidence and success. Other studies in Malaysia have applied different predictive models to mathematics performance but have mostly focused on face-to-face learning. For example, Lye *et al.* (2010) applied various ML techniques including Back-propagation Neural Networks (BPNN), Classification and Regression Trees (CART) and Generalized Regression Neural Networks (GRNN) using enrollment and exam performance data from pre-university students to predict their mid-semester and final exam scores in mathematics. Among these, BPNN achieved the highest prediction accuracy. Similarly, Samsudin *et al.* (2022) utilized SVM regression on undergraduate CGPA data and found that the radial basis function kernel produced the most accurate academic performance predictions.

In addition, Suthar and Tarmizi (2010) applied logistic regression to survey data from university undergraduates, showing that students' beliefs about mathematics and their self-confidence are important predictors of academic success. Yahaya and Hasan (2021) used absorbing Markov chain models on enrollment and academic performance data of undergraduate mathematics students to analyze patterns of retention, progression, and graduation, demonstrating how probabilistic models can provide clear insights into academic trajectories. More broadly, research has highlighted the impact of advanced technologies and psychological factors in mathematics education. For instance, Jin *et al.* (2022) reviewed how Industry 4.0 technologies like Big Data and Artificial Intelligence could transform teaching practices and decision-making in Malaysian mathematics education.

Overall, BNs in online mathematics education have proven effective at utilizing detailed, step-by-step problem-solving data and capturing the causal relationships between different knowledge components. This enables the delivery of personalized feedback and the creation of adaptive learning pathways. However, in Malaysia, research applying predictive modelling, especially BNs to mathematics learning in higher education is still limited and tends to focus mainly on traditional face-to-face settings rather than online or e-learning environments. Despite the rapid growth of digital education platforms, Malaysian studies have yet to fully explore the use of predictive BN models that integrate comprehensive, real-time learning analytics data. This study seeks to address this gap by developing a BN model designed to predict student performance in online mathematics courses using rich learning analytics data. Such a model aims to provide a deeper and more nuanced understanding of student learning behaviors, supporting personalized interventions tailored to the specific needs of online learning in Malaysian higher education.

3. Bayesian Network Framework for Performance Modelling

In this section, we present a framework for developing a BN model to analyze student performance using learning analytics. Figure 1 illustrates the framework, which integrates learning analytics principles with BN modeling.

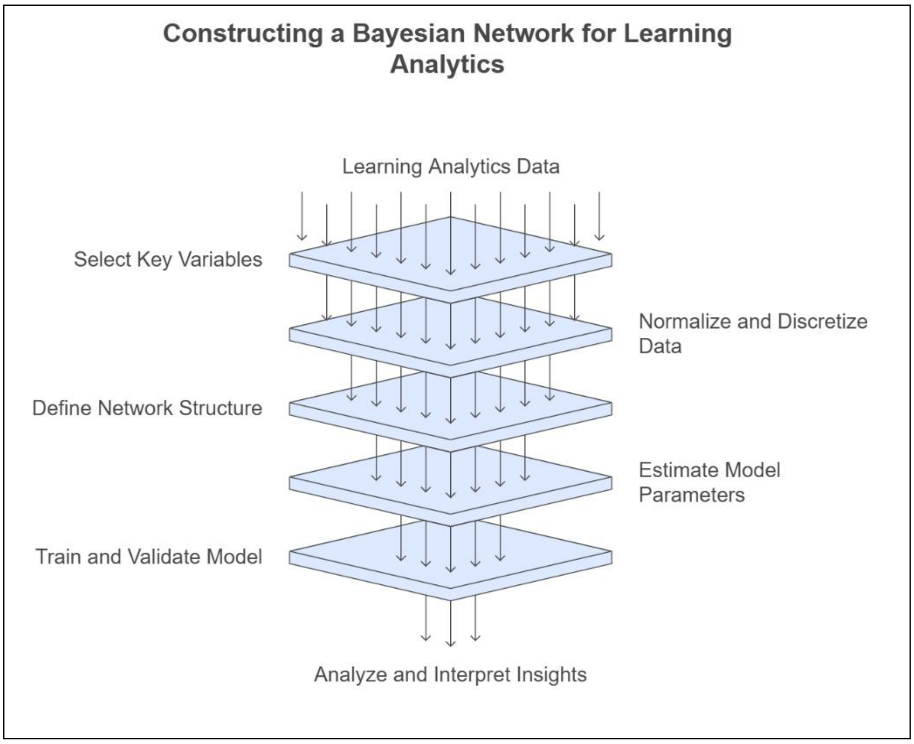

Figure 1: Framework for constructing Bayesian Network for learning analytics

The detailed process can be explained as following steps:

(1) Select key variables. The first step in building a BN is to identify and select relevant learning analytics variables that significantly impact student performance. These variables serve as nodes in the network and can be categorized into academic (e.g. GPA, exam scores and assignment results) and behavioral data (e.g. log activity, engagement levels, and interaction with learning resources). The selection of these variables is based on their relevance, availability, and predictive power in determining learning outcomes.

(2) Normalize and discretize data. This step is data preprocessing, which ensures consistency and compatibility with BN modeling. Normalization is used to transform numerical values into a standardized range, allowing for better comparability across different data points. Discretization converts continuous variables into categorical states which makes the data more suitable for Bayesian inference.

(3) Define network structure. The network structure defines the causal relationships between variables, ensuring a logical and interpretable model. This can be done either data-driven structure learning or expert-driven structure, or combination of both approaches. The structure must also respect chronological order, ensuring dependencies reflect realistic learning progressions (e.g., prior coursework influencing final grades).

(4) Estimate model parameters. Once the BN structure is established, the next step is to estimate its parameters by determining the prior probabilities of parent nodes and the CPTs for child nodes. CPTs define the likelihood of a node taking a specific value based on the values of its parent nodes which captures the relationships between variables. These probabilities can be estimated using methods such as MLE, Bayesian estimation, or expert elicitation.
(5) Train and validate model. To evaluate its performance, the dataset is divided into training and test sets, ensuring that the model learns from one portion of the data and is tested on another. Cross-validation techniques, such as k-fold validation, are used to assess the model's reliability and consistency. Additionally, the predictions made by the BN are compared with traditional ML models to determine its effectiveness.
(6) Analyze and interpret insights. Once the BN is trained, it provides valuable insights into student performance patterns by analyzing causal relationships between student behaviors and academic outcomes. The model also enables early identification of at-risk students by examining probabilistic dependencies, allowing for timely interventions. Educators and administrators can use these insights to identify key factors influencing academic success and helping them prioritize strategies that have the most significant impact on student outcomes.

Regarding steps (5) and (6), some researchers prefer to first analyze and interpret model insights, while others prioritize training and validation. This interchangeability reflects the iterative nature of model development, where initial analyses inform subsequent validation and refinement. In the following section, we present the process of applying the BN for predictive analysis within the proposed framework.

4. Application of the Proposed Framework

This section explains how the proposed BN framework is applied to predict student performance in solving system of linear equations using Mathematica software. We begin by defining the dataset and key learning analytics variables, followed by developing the BN including its structure and probability estimation. Finally, we evaluate the model's accuracy and reliability using performance metrics.

4.1. *Dataset and variables*

In this study, data were collected from students enrolled in the Mathematics Software Laboratory course at the School of Distance Education (SDE), Universiti Sains Malaysia (USM), during the 2023–2024 academic session. The dataset comprises educational records of nine students enrolled in the course during this period. SDE conducts online learning through the Moodle LMS, utilizing a blended learning approach. Students are required to attend online lectures via the Webex platform, engage in independent learning activities and complete various assessments integrated within the Moodle LMS, including multiple-choice quizzes, short-answer exercises, and problem-solving assignments designed to evaluate their understanding of mathematical concepts and software applications. To simplify BN modeling, we focus on data from a specific subtopic in the course—solving systems of linear equations using Mathematica software.

Table 1: The variables used for the construction of the Bayesian Network (BN) model

Variable (nodes)	Description	States
(i) Student engagement		
Course_hits	The number of times a student accesses the course materials in the LMS.	Low, Medium, High
Days_with_access	The total number of days a student logs into the course platform.	Low, Medium, High
Resources_with_access	The amount of learning resources a student interacts with.	Low, Medium, High
Course_Participation	Measures a student's engagement in activities in the LMS.	Low, High
Attendance_Rate	The percentage of scheduled classes or sessions a student attends.	Low, High
(ii) Resource untilization		
Familiarity_with_Mathematica	Indicates whether a student has prior experience using Mathematica	Yes, No
Course_Module_Basic_Functions	Completion of Basic Functions module	Complete, Not complete
Course_Module_SLE	Completion of System of Linear Equations (SLE) module	Complete, Not complete
Webex_Lecture_Module_Basic_Functions	Attending Webex lecture on Basic Functions.	Complete, Not complete
Video_Mathematica_for_Beginner	Accessed the introductory Mathematica video	Complete, Not complete
Webex_Lecture_Module_SLE	Attended the Webex lecture on solving SLE.	Complete, Not complete
Video_Matrix_Manipulation	Accessed the video on matrix manipulation.	Complete, Not complete
(iii) Self-Learning and assessment		
Self_Learning_Basic_Calculation	Engaged in self-learning activities for basic calculations.	Complete, Not complete
Self_Learning_Matrices	Engaged in self-learning activities related to matrices.	Complete, Not complete
Gamification_Quiz_Basic_Calculation	Participation in gamified quizzes on basic calculations.	Complete, Not complete
Assessment_Solving_SLE	Completion of the assessment on solving SLE.	Complete, Not complete
(iv) Performance		
Mathematica_Laboratory_Exercise	Participation in lab exercises using Mathematica.	Low, Medium, High
Examination_Grade	The final exam score categorized into different performance levels.	Medium, High
Solving_SLE	Achievement in solving the system of linear equations.	Low, Medium, High

The BN model captures the relationships among 19 learning analytics variables grouped into several key domains. The BN model captures the relationships among 19 learning analytics variables grouped into several key domains. These include measures of student engagement (e.g., frequency of course access, participation levels and attendance rates); resource utilization variables, reflecting students' interactions with various instructional materials (e.g., online lectures, course modules and tutorial videos); self-directed learning activities and assessment variables (e.g., use of open resources and online gamification); and academic performance indicators (e.g., laboratory exercises, examinations and problem-solving tasks) to comprehensively evaluate learning outcomes within the topic. The variables used for the construction of BN model are listed in Table 1. Some variables are discretized into categories to represent distinct states in the BN. For example, "Course_hits", "Days_with_access" and "Resources_with_access" are discretized into three categories Low, Medium, and High to better represent engagement levels. Meanwhile, "Familiarity_with_Mathematica" is categorized as Yes or No to indicate whether a student has prior experience using the software. The variables "Course_Module_Basic_Functions" and "Course_Module_SLE" are defined by completion status, classified as either Complete or Not Complete.

4.2. *BN modelling of student performance*

The process of fitting a BN is called 'learning'. In this process we are looking for an optimal configuration for our model that fully describes the relationship between the variables (the dependency structure of the variables). An optimal configuration must be well-suited for its corresponding application. With this regard, there are two approaches for designing the configurations for BN, leveraging expert domain-knowledge and utilizing mathematical algorithms. Given the limitations of available data and the need to include pedagogical knowledge, we utilized expert elicitation to define the network structure. Specifically, pedagogical experts were consulted to identify and define the relationships among key variables of student engagement, resource utilization, self-directed learning activities and assessment, and academic performance. Although expert-driven network construction can introduce potential biases, preliminary experiments demonstrated the potential for strong predictive performance.

In this study, the network structure was developed through elicitation with a domain expert—a mathematics lecturer who is well-versed in the subject matter relevant to the model. The elicitation was conducted through a structured verbal interview one-on-one format allowed for in-depth exploration of the expert's knowledge and reasoning, enabling for clarify any ambiguities and probe specific relationships among variables systematically. The conversation was recorded and transcribed, and key elements shaping the network structure were extracted by identifying consistent causal links and dependencies mentioned by the expert. This approach aligns with practices in similar expert-elicited BN modeling studies, such as the work by Marcot *et al.* (2006), where a single expert or a small group of experts provide detailed knowledge through structured interviews, which is then carefully translated into the BN structure.

The constructed BN of the learning analytics variables and student performance in solving system of linear equations using Mathematica software is shown in Figure 2. It represents the relationships between various learning analytics metrics and student activities in an online mathematics course. The input nodes represent student engagement indicators, including course hits, days with access, resource availability, familiarity with Mathematica, attendance rate, and course participation. These factors influence student engagement with different learning modules, such as Webex lectures, video tutorials, self-learning exercises, and quizzes. These resources act as intermediaries, helping students build foundational knowledge before they attempt assessments. The final node, Solving System of Linear Equation (SLE) using Mathematica software, represents a key competency in the course and is the outcome of a sequence of learning interactions.

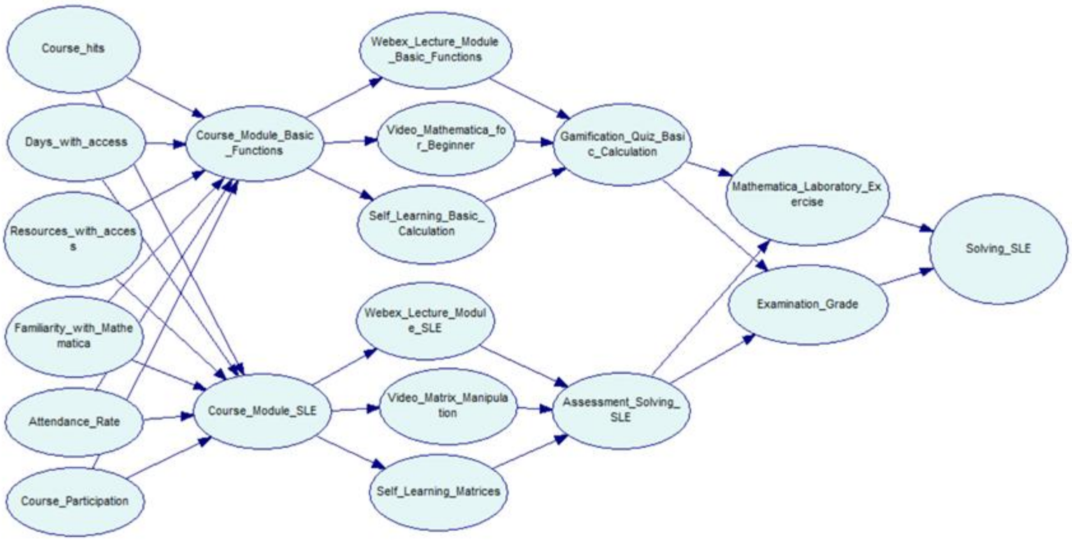

Figure 2: The constructed Bayesian Network (BN) of learning analytics indicator and student performance in solving systems of linear equations using Mathematica software

Once the model configuration is determined, the parameters are estimated using the data in the training set. Specifically, the Expectation-Maximization (EM) algorithm was employed to learn the parameters, iteratively updating the CPTs tables based on observed data. The EM algorithm using the fundamental probabilistic principles of BN using Eq. (1) and Eq. (2) to refine CPTs through alternating expectation and maximization steps. This approach was particularly suitable for handling potential missing data within our dataset, a common challenge in educational learning analytics. The BN modeled in GeNIe Modeler (see Figure 3) visually represents the network structure and the learned probabilities.

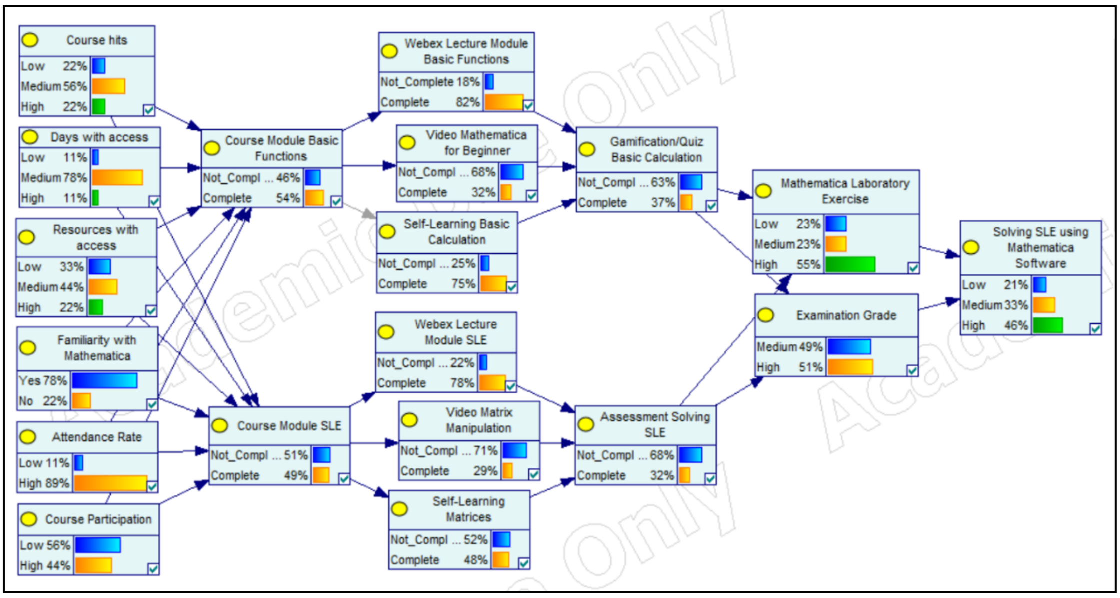

Figure 3: Bayesian Network (BN) utilizing Expectation-Maximization (EM) for parameter learning

This BN model in Figure 3 indicates a causal chain of factors culminating in varying levels of proficiency in "Solving SLE using Mathematica Software." High "Attendance Rate" (89%) and "Familiarity with Mathematica" (78% Yes) provide a strong foundation, leading to substantial engagement with initial modules like "Webex Lecture Module Basic Functions" (82% Complete) and "Course Module Basic Functions" (54% Complete). However, a significant decline in completion rates for "Video Mathematica for Beginner" (33% Complete) and "Gamification/Quiz Basic Calculation" (37% Complete) suggests an early struggle with applying basic concepts. This challenge persists in advanced activities like "Video Matrix Manipulation" (29% Complete) and "Assessment Solving SLE" (32% Complete), indicating a growing difficulty in translating theoretical knowledge into practical application. Despite a considerable portion of students achieving "Medium" examination grades (49%), this does not translate into high proficiency in the practical application of solving SLE using Mathematica, where only 46% achieve a "High" score. The network reveals that high engagement at each stage significantly increases the likelihood of success in the subsequent learning activities, culminating in a higher probability of achieving high proficiency in solving SLE using Mathematica, though a significant portion of students still demonstrate low proficiency (21%), indicating potential gaps in the learning process despite strong foundational elements.

One of the most powerful capabilities of BN is their ability to perform inference. As illustrated in Step (6) (see Figure 1), interpreting insights provides valuable understanding of performance patterns, including the early identification of at-risk students through probabilistic dependencies. In this study, a diagnostic inference mechanism is employed to identify students struggling to master systems of linear equations in Mathematica and to uncover the underlying factors contributing to their difficulties. Applying Eq. (3), we observed evidence of 100% of students scoring low in solving SLE using Mathematica, resulting in the subsequent update of all other node states within the network (see Figure 4). One of the most powerful capabilities of Bayesian Networks is their ability to perform inference. Diagnostic inference, in particular, allows for the identification of high-risk students struggling with mastering systems of linear equations in Mathematica and helps reveal the factors behind their difficulties.

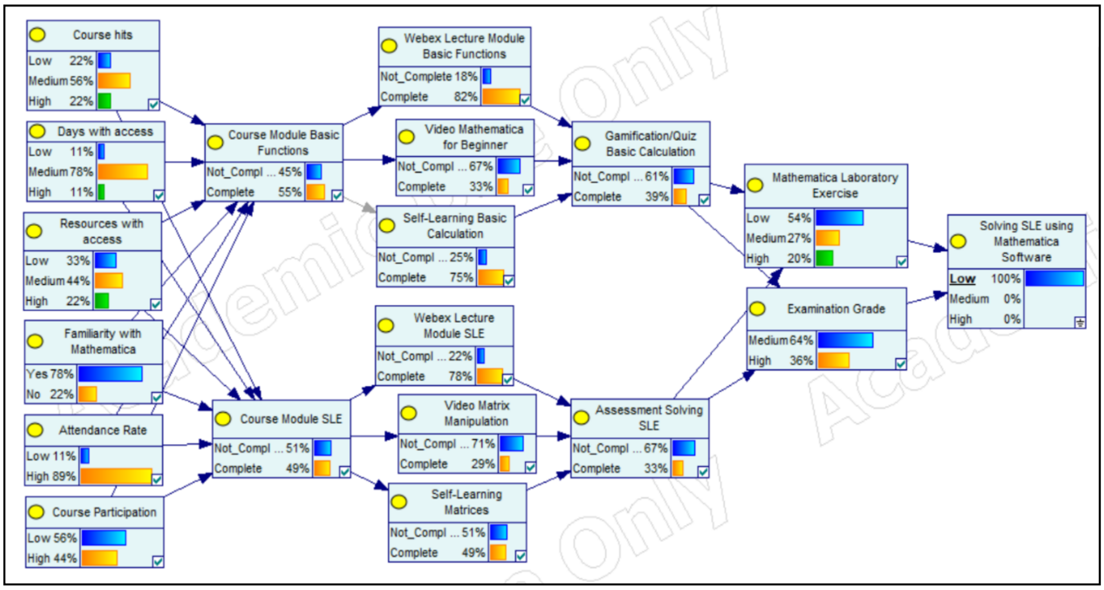

Figure 4: Diagnostic inference in Bayesian Network (BN) for identifying key learning challenges

This BN model reveals a 100% low score in "Solving SLE using Mathematica Software," highlights a significant disconnect between initial engagement and ultimate performance. Despite a strong foundation indicated by high "Attendance Rate" (89%) and "Familiarity with Mathematica" (78% Yes), students encounter critical challenges in applying their knowledge to complex tasks. The initial engagement with "Course Module Basic Functions" (55% Complete) and "Webex Lecture Module Basic Functions" (82% Complete) suggests a willingness to learn. However, a sharp decline in engagement with "Video Mathematica for Beginner" (33% Complete) marks the first major red flag, indicating a potential knowledge gap early on. This decline continues with "Video Matrix Manipulation" (29% Complete) and "Assessment Solving SLE" (33% Complete), demonstrating a struggle to translate theoretical understanding into practical problem-solving. While "Examination Grade" shows a significant portion achieving medium grades (64%), this knowledge does not translate into practical application, as evidenced by the 100% low score in the culminating task. This network suggests that while students are present and possess foundational knowledge, they struggle with advanced concepts and practical application, indicating a need for targeted interventions focusing on bridging the gap between theory and practice.

In the next section, the result of applying this model for predicting the student performance is presented.

4.3. *Model evaluation*

BNs are typically evaluated using classification metrics such as accuracy, precision, recall, and F1-score to assess their predictive performance (Heckerman 1997; Korb & Nicholson 2008). Accuracy measures the overall correctness of predictions, precision quantifies the proportion of true positive predictions among all positive predictions, recall (or sensitivity) measures the proportion of true positives detected among all actual positives, and F1-score provides a harmonic mean balancing precision and recall. Most of these metrics are derived from the confusion matrix, which cross-tabulates the counts of correctly and incorrectly predicted instances for each class. In a binary classification model, the confusion matrix consists of four key components: True Positives (TP), representing the number of correctly predicted positive cases; True Negatives (TN), representing the number of correctly predicted negative cases; False Positives (FP), representing negative cases incorrectly predicted as positive; and False Negatives (FN), representing positive cases incorrectly predicted as negative. Essentially, the diagonal cells (from the top-left to the bottom-right) contain the correctly classified instances, while all off-diagonal cells represent misclassifications. The following performance metrics are used to evaluate the prediction methods:

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{4}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

$$\text{F1-Score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{7}$$

Table 2: Performance measures of BN model

Metric	Value	Mean ± Standard Deviation (SD)
Accuracy	0.7778	(0.4556 ± 0.366)
Precision	0.8413	(0.4191 ± 0.3886)
Recall	0.7778	(0.4556 ± 0.366)
F1	0.7407	(0.4168 ± 0.3724)

To evaluate the BN model's performance, we used a 70/30 train-test split where 70% of the data was used for training and the remaining 30% was reserved for testing to assess predictive accuracy. Table 2 shows the results of the performance BN model. The model achieved an accuracy of 0.7778, indicating that approximately 78% of the predictions were correct. However, the relatively large standard deviation (±0.366) in accuracy suggests some variability across different test samples or folds, implying that the model's consistency may fluctuate depending on the data subset. Precision was notably high at 0.8413, demonstrating that when the model predicts a positive outcome, it is correct over 84% of the time. This high precision, despite a moderate standard deviation (±0.3886), reflects the model's strength in minimizing false positive predictions. The recall value matched the accuracy at 0.7778, which means the model successfully identified about 78% of the actual positive cases, though the standard deviation again indicates variability in sensitivity across samples. The F1-score, which balances precision and recall, was 0.7407, suggesting a reasonably good overall balance between correctly identifying positive cases and limiting false positives. The associated standard deviation (±0.3724) underscores some instability in performance but still supports the model's robustness. Overall, these results indicate that the developed BN model performs well in predicting the target classes, with strong precision and balanced recall. The observed variability highlights the need for further validation and potential model refinement to ensure consistent performance across diverse dataset.

5. Conclusion

The present study introduced a BN framework utilizing learning analytics indicators to predict student performance in online learning. To achieve this, the framework modeled academic success through a sequence of six key steps for constructing a BN tailored for this predictive task. These steps encompassed: (1) the selection of pertinent academic and behavioral variables; (2) the normalization and discretization of the data; (3) the definition of the network structure to represent underlying causal relationships; (4) the estimation of the model parameters; (5) the training and validation of the model's predictive capabilities; and (6) the analysis and interpretation of the resulting insights to elucidate performance patterns and facilitate the identification of students at potential academic risk, thereby enabling timely interventions. The proposed framework was subsequently applied to student data obtained from individuals enrolled in an online Mathematics subject at public universities in Malaysia. The constructed BN model revealed critical insights into the dynamics of student learning pattern, notably indicating a direct correlation between low student engagement in core learning activities and diminished competency in achieving key course outcomes. This observation was further confirmed by the model's inference mechanism, which highlighted the significant negative impact of low engagement on overall student performance.

A primary contribution of this work is the distinctive approach in developing BN in online education, achieved by integrating learning analytics and mapping the subject's learning path. This approach allows for accurate student performance prediction, thereby enabling the early detection of at-risk students and the subsequent implementation of targeted interventions.

Future research should prioritize the enhancement of model accuracy by incorporating temporal factors, such as the timing of prerequisite courses, alongside personal, social, and psychological variables. This would provide a more comprehensive understanding of student capacity. Ultimately, this study demonstrates the potential of BN powered by learning analytics to contribute meaningfully to the improvement of educational practices and student outcomes within online learning settings.

Acknowledgments

We sincerely appreciate the generous support of R502 - KR - ARU005 - 0000000588 - K134 GGPM-2022-040 (Geran Penyelidikan Akademik 2024) from Universiti Sains Malaysia in funding this research work.

References

Ahmed E. 2024. Student performance prediction using machine learning algorithms. *Applied Computational Intelligence and Soft Computing* **2024**: 4067721.

Almalawi A, Soh B., Li A. & Samra H. 2024. Predictive models for educational purposes: A systematic review. *Big Data and Cognitive Computing* **8**(12): 187.

Alshahrani A. 2023. Revolutionizing blended learning: Exploring current trends and future research directions in the era of ChatGPT. *Proceedings of 2023 7th International Conference on Business and Information Management*, pp. 41-47.

Arnold K.E. & Pistilli M.D. 2012. Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, pp. 267-270.

Baker R.S. & Inventado P.S. 2014. Educational data mining and learning analytics. In Larusson J.A. & White B. (eds.). *Learning analytics: From research to practice*: 61–75. New York, NY: Springer.

BayesFusion LLC. 2023. GeNIe Modeler [software]. https://www.bayesfusion.com/genie/

Bellarhmouch Y., Majjate H., Jeghal A., Tairi H. & Benjelloun N. 2025. Detecting student engagement in an online learning environment using a machine learning algorithm. *Informatics* **12**(2): 44.

Bird K.A., Castleman B.L., Mabel Z. & Song Y. 2021. Bringing transparency to predictive analytics: A systematic comparison of predictive modeling methods in higher education. *AERA Open* **7**(1): 1-19.

Brinton C.G., Buccapatnam S., Chiang M. & Poor H.V. 2016. Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance. *IEEE Transactions on Signal Processing* **64**(14): 3677-3692.

Chen F. & Cui Y. 2020. Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics* **7**(2): 1-17.

Chen X., Zong X., Zhang D. & Cheng S. 2024. Evaluation of teaching effectiveness based on Bayesian network algorithm in teaching and learning process in higher education institutions. *Journal of Electrical Systems* **20**(6): 1800-1810.

Conati C., Gertner A. & VanLehn K. 2002. Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction* **12**(4): 371–417.

Culbertson M.J. 2016. Bayesian networks in educational assessment: The state of the field. *Applied Psychological Measurement* **40**(1): 3-21.

Daniel B. 2015. Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology* **46**(5): 904-920.

Delen D., Topuz K. & Eryarsoy E. 2020. Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research* **281**(3): 575-587.

Fan Y., Zhang J., Zu D. & Zhang H. 2021. An automatic optimal course recommendation method for online math education platforms based on Bayesian model. *International Journal of Emerging Technologies in Learning* **16**(13): 95–107.

Ferguson R. 2012. Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning* **4**(5-6): 304-317.

Fujiwara Y. 2016. Predicting student dropout using machine learning techniques. *Proceedings of the 9th International Conference on Educational Data Mining*, pp. 588-591.

Habti F.E.E., Hiri M., Chrayah M., Bouzidi A. & Aknin N. 2025. Enhancing student performance prediction in e-learning ecosystems using machine learning techniques. *International Journal of Information and Education Technology* **15**(2): 301-311.

Heckerman D. 1997. Bayesian networks for data mining. *Data Mining and Knowledge Discovery* **1**: 79-119.

Jayaprakash S.M., Moody E.W., Lauría E.J.M., Regan J.R. & Baron J.D. 2014. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics* **1**(1): 6-47.
Jensen F.V. & Nielsen T.D. 2007. *Bayesian Networks and Decision Graphs.* 2nd Ed. New York, NY: Springer Science & Business Media.
Jiang S., Huang X., Sung S.H. & Xie C. 2023. Learning analytics for assessing hands-on laboratory skills in science classrooms using Bayesian network analysis. *Research in Science Education* **53**(2): 425-444.
Jin S.J., Abdullah A.H., Mokhtar M. & Abdul Kohar U.H. 2022. The potential of big data application in mathematics education in malaysia. *Sustainability* **14**(21): 13725.
Jose B.C., Kumar M.A., UdayaBanu T. & Nagalakshmi M. 2024. Assessing the effectiveness of adaptive learning systems in K-12 education. *International Journal of Advanced IT Research and Development* **1**(1): 1-8.
Käser T., Klingler S., Schwing A.G. & Gross M. 2017. Dynamic Bayesian networks for student modeling. *IEEE Transactions on Learning Technologies* **10**(4): 450-462.
Kato Nabirye H. 2025. E-learning management systems: Best practices for implementation. *Research Invention Journal of Current Issues in Arts and Management* **4**(2):43-47.
Kim B.H., Vizitei E. & Ganapathi V. 2018. GritNet: Student performance prediction with deep learning. *Proceedings of the 11th International Conference on Educational Data Mining*, pp. 1–5.
Kitson N.K., Constantinou A.C., Guo, Z., Liu Y. & Chobtham K. 2023. A survey of Bayesian network structure learning. *Artificial Intelligence Review* **56**: 8721-8814.
Koller D. & Friedman N. 2009. Probabilistic graphical models: Principles and techniques. Cambridge, MA: MIT Press.
Kondo N. & Hatanaka T. 2018. An approach for estimation of students' learning states using Bayesian network with LMS Log Data. *7th International Congress on Advanced Applied Informatics,* pp. 474-477.
Kondo N. & Hatanaka T. 2019. Estimation of students' learning states using Bayesian networks and log data of learning management system. *International Journal of Institutional Research and Management* **3**(2): 35-49.
Korb K.B. & Nicholson A.E. 2008. The causal interpretation of Bayesian networks. In Holmes D.E. & Jain L.C. (eds.). *Innovations in Bayesian Networks: Theory and Applications*: 83-116. Heidelberg: Springer Berlin.
Lazaro G.Rd. & Duart J.M. 2023. Moving Learning: A Systematic Review of Mobile Learning Applications for Online Higher Education. Journal of New Approaches in Educational Research **12:** 198–224.
Lee C.A., Tzeng J.W., Huang N.F. & Su Y.S. 2021. Prediction of student performance in massive open online courses using deep learning system based on learning behaviors. *Educational Technology & Society* **24**(3): 130-146.
Li G., Cui, J., Fu, H. & Sun, Y. 2024. Light GBM and GA based algorithm for predicting and improving students' performance in higher education in the context of big data. *7th International Conference on Education, Network and Information Technology*, pp. 1-5.
Liu F., Fan Z., Huang F., Li Y., He Y. & Hu W. 2022. Modeling Learner Behavior Analysis Based on Educational Big Data and Dynamic Bayesian Network. *IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 48-53.
Looi Z.N., Song P.C., Lim H.T. & Looi S.Y. 2023. A case study via Bayesian network: Investigating factors influencing student academic performance in online teaching and learning during COVID-19 pandemic. *The International Conference on Data Science and Emerging Technologies*, pp. 303–317.
Lu O.H.T., Huang A.Y.Q., Huang J.C.H., Lin A.J.Q., Ogata H. & Yang S.J.H. 2018. Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society* **21**(2): 220–232.
Luan H. & Tsai C.C. 2021. A review of using machine learning approaches for precision education. *Educational Technology & Society* **24**(1): 250-266.
Lye C.T., Ng L.N., Hassan M.D., Goh W.W., Law C.Y., Ismail N. 2010. Predicting pre-university students' Mathematics achievement. *Procedia - Social and Behavioral Sciences* **8**: 299–306.
Lyn J.T.Y., Kheng G.Y., Chow L.A. & Meng N.Y. 2024. Learning Analytic Framework for students'academic performance and critical learning pathways. *Journal of Quality Measurement and Analysis JQMA* **20**(2): 127-147.
Millán E., Descalço L., Castillo G., Oliveira P. & Diogo S. 2013. Using Bayesian networks to improve knowledge assessment. *Computers & Education* **60**(1): 436-447.
Millán E., Loboda T. & Pérez-de-la-Cruz J.L. 2010. Bayesian networks for student model engineering. *Computers & Education* **55**(4): 1663-1683.
Moreno-Marcos P.M., Pong T.C., Muñoz-Merino P.J. & Delgado Kloos C. 2020. Analysis of the factors influencing learners' performance prediction with learning analytics. *IEEE Access* **8**: 5264-5282.
Murphy K.P. 2002. Dynamic Bayesian networks: representation, inference and learning. PhD Thesis. Berkeley: University of California.
Namoun A. & Alshanqiti A. 2021. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences* **11**(1): 237.

Nespereira C.G., Elhariri E., El-Bendary N., Vilas A.F., Redondo R.P.D. 2016. Machine learning based classification approach for predicting students performance in blended learning. *The 1st International Conference on Advanced Intelligent System and Informatics*, pp. 47-56.
Ong H.C. & Lim J.S. 2014. Identifying factors influencing mathematical problem solving among matriculation students in Penang. *Pertanika Journal of Social Science and Humanities* **22**(3): 393-408.
Ouyang F., Wu M., Zheng L. & Jiao P. 2023. Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education* **20**(1): 4.
Pakdaman M., Moghadam M.N., Dehghan H.R., Dehghani A. & Namayandeh M. 2019. Evaluation of the Cost-Effectiveness of Virtual and Traditional Education Models in Higher Education: A Systematic Review. *Health Tech Asmnt Act.* **3**(1): e5715.
Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
Pihlajamaa J., Karukka M. & Ålander H. 2016. Comparison of higher education student and teacher perceptions of e-learning. *Proceedings of the 15th European Conference on e-Learning*, pp. 782-785.
Qiu W., Khong A.W.H, Supraja S. & Tang W. 2024. A dual-mode grade prediction architecture for identifying at-risk students. *IEEE Transactions on Learning Technologies* **17**: 803-814.
Rizwan S., Nee C.K. & Garfan S. 2025. Identifying the factors affecting student academic performance and engagement prediction in MOOC using deep learning: A systematic literature review. *IEEE Access* **13**: 18952-18982.
Romero C. & Ventura S. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* **33**(1): 135-146.
Russell S.J. & Norvig P. 2021. *Artificial Intelligence: A Modern Approach*. 4th Ed. Harlow, UK: Pearson Education Limited.
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.
Seffrin H., Bittencourt I.I., Isotani S. & Jaques P.A. 2016. Modelling students' algebraic knowledge with dynamic Bayesian networks. *IEEE 16th International Conference on Advanced Learning Technologies*, pp. 44-48.
Siemens G. 2013. Learning analytics: The emergence of a discipline. *American Behavioral Scientist* **57**(10): 1380-1400.
Siemens G. & Baker R.S. 2012. Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* pp. 252-254.
Siemens G. & Long P. 2011. Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review* **46**(5): 30-32.
Sulak S.A. & Koklu N. 2024. Predicting student dropout using machine learning algorithms. *Intelligent Methods in Engineering Sciences* **3**(3): 91-98.
Suthar V. & Tarmizi R.A. 2010. Effects of students' beliefs on Mathematics and achievement of university students: Regression analysis approach. *Journal of Social Sciences* **6**(2): 146-152.
Wan H., Li M., Zhong Z. & Luo X. 2023. Early prediction of student performance with LSTM-based deep neural network. *IEEE 47th Annual Computers, Software, and Applications Conference*, pp. 132-141.
Wang J. & Yu Y. 2025. Machine learning approach to student performance prediction of online learning. *PLoS ONE* **20**(1): e0299018.
Wang C. & Han D. 2016. Data mining technology based on Bayesian network structure applied in learning. *Advanced Science and Technology Letters* **137**: 67–71.
Wong B.T.M., Li K.C. & Liu M. 2025. The role of learning analytics in evaluating course effectiveness. *Sustainability* **17**(2): 559.
Yahaya K.H. & Hasan H. 2021. Application of Markov chain in students' assessment and performance: A case study of School of Mathematical Sciences, one of the public university in Malaysia. *Proceedings of the 16th IMT-GT International Conference on Mathematics, Statistics and their Applications* **36**: 01004.
Zhao Z., Kang F., Wang J., Chen B., Yang M. & Qu S. 2023. Analysis and prediction of the factors influencing students' grades based on their learning behaviours in MOOCs. *Proceedings of The International Conference on Computer Science and Education*, pp. 355-368.
Zheng C. 2025. Application of Bayesian networks in adaptive listening assessment system in flipped English learning environment. *Journal of Computational Methods in Sciences and Engineering* **25**(4): 3197-3209.

School of Distance Education
Universiti Sains Malaysia
11800 USM Penang, MALAYSIA
*E-mail: rnurulhuda@usm.my**

School of Mathematical Sciences
Universiti Sains Malaysia
11800 USM Penang, MALAYSIA
E-mail: m.tahir@usm.my

Received: 20 March 2025
Accepted: 16 June 2025

*Corresponding author