I.J. Modern Education and Computer Science, **2026, 1, 21-47**
Published Online on February 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2026.01.02

Developing a Bayesian Network Model to Predict Students' Performance Based on the Analysis of their Higher Education Trajectory

Thabo Ramaano*
School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa
E-mail: 1134539@students.wits.ac.za
ORCID iD: https://orcid.org/0000-0002-5484-7149
*Corresponding Author

Ashwini Jadhav
Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa
E-mail: ashwini.jadhav@wits.ac.za
ORCID iD: https://orcid.org/0009-0000-9040-6562

Ritesh Ajoodha
School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa
E-mail: ritesh.ajoodha@wits.ac.za
ORCID iD: https://orcid.org/0000-0002-6443-8592

Received: 16 January, 2025; Revised: 21 June, 2025; Accepted: 27 October, 2025; Published: 08 February, 2026

Abstract: The Admission Point Score (APS) metric, commonly used to admit prospective students into academic programmes, may appear effective in predicting student success. In reality, almost 50% of students admitted to a science programme in a higher education institution failed to meet all the requirements necessary to complete the programme during the period of 2008 and 2015. This had a direct impact on the overall graduation throughput. This research therefore focuses on adopting a probabilistic-graphical approach as a viable alternative to the APS metric for determining student success trajectories in higher education. The purpose of this approach was to provide higher education institutions with a system to monitor students' academic performance en-route to graduation from a probabilistic and graphical point of view. This research employed a probability distribution distance metric to ascertain how close the learned models were to the true model for varying sample sizes. The significance of these results addressed the need for knowledge discovery of dependencies that existed between the students' module results in a higher education trajectory that spans three years.

Index Terms: Bayesian Network, Higher Education Trajectory, Hill-Climbing Structure Learning Algorithm, Kullback-Leibler Divergence, Variable Elimination Inference Algorithm

1. Introduction

South African higher education institutions are facing growing concerns over high dropout rates and low graduation throughputs amongst students [1]. According to a report on higher education and skills in South Africa conducted by Statistics South Africa, it has been established that the on-time graduation rate for students pursuing three-year degree programmes across all higher education institutions in South Africa stood at less than 30% [2].

The Admission Point Score (APS) metric is currently adopted by South African higher education institutions as an admission indicator to admit prospective students. As a result, this admission metric has not yielded the desired results of achieving a plausible graduation throughput.

According to Fig.1. almost 50% of students admitted for degree studies failed to complete all the necessary requirements for their respective academic courses when considering medium risk and high risk students. To cater for this, an improved mechanism is required to supplement the current APS metric with the purpose of improving the overall on-time graduation throughput.

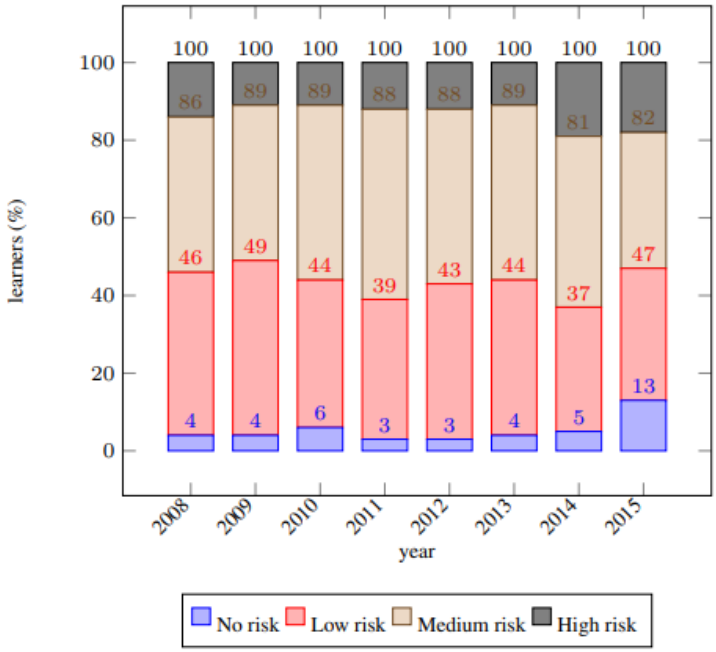

Fig. 1. Percentage of students (categorised as no-risk; low-risk; medium-risk and high-risk) registered at a higher education institution between 2008 and 2015 adopted from [3].

A student's higher-education trajectory refers to the academic journey undertaken to obtain a qualification within the prescribed period. As such, the trajectory can be deemed a success (student has obtained a qualification) or failure (student has discontinued studies). In simpler terms, it can be thought of as the progression through the higher education pipeline [4]. The illustration in Fig.2. is a graphical representation of a higher education academic trajectory that spans three years and depicts both the success and failure that emanates in a student's higher education trajectory.

According to [4], the topic of academic trajectory of students in higher education, en-route to obtaining their academic qualifications, is an under-researched issue despite its relevance in determining student success. In addition, [4] provided the illustration in Fig.3. that depicts the three levels of predictors, namely: macro-level, meso-level and micro-level in the higher education domain. The aforementioned higher education predictors have the potential of shaping a student's higher education trajectory. This study will make use of academic performances to determine student success trajectories which is an attribute located in the micro-level of predictors.

It has been well documented that higher educational institutions are looking to improve student graduation throughputs and reduce dropout rates. The academic success of students is reliant on higher education institutions administering support programmes to improve their on-time completion rates and reduce dropout rates. The emergence of Educational Data Mining techniques has made it possible to use student data to find student attributes that have a significant impact on predicting academic performance and identify learning outcomes [5]. The analysis involves identifying behavioural and learning patterns and implements recommendation systems for students. Machine learning has made it possible to predict the performance of students thus finding good predictors of performance.

A number of studies have made significant advancements in the domain of predicting student performance using a variety of predictive algorithms. Results of particular interest are those emanating from studies using probabilistic graphical models [6,7,8]. The aforementioned studies provide useful methodologies and concepts required for this research. These include: the construction of a Bayesian model [7]; real world application of using a Bayesian model in an education environment [6] and taking advantage of the graphical nature of Bayesian modelling to get a deeper understanding of the relationships that influence the success of a student [8].

The limitations of the aforementioned predictive algorithms lie in their inability to explore the students' higher education trajectory from first year to final year. An analysis of this nature may provide researchers and educators with valuable insights to optimize the probability of on-time graduation by monitoring students' higher education academic journey.

The purpose of this research is to employ a Bayesian probabilistic approach to predict a student's higher education trajectory and provide higher education institutions with a simplified probabilistic-graphical view for admitting new students. A model of this nature is useful in assisting higher education institutions monitor students' academic trajectory with the aim of employing guidance tools that will assist them in ensuring that students complete their degrees on time.

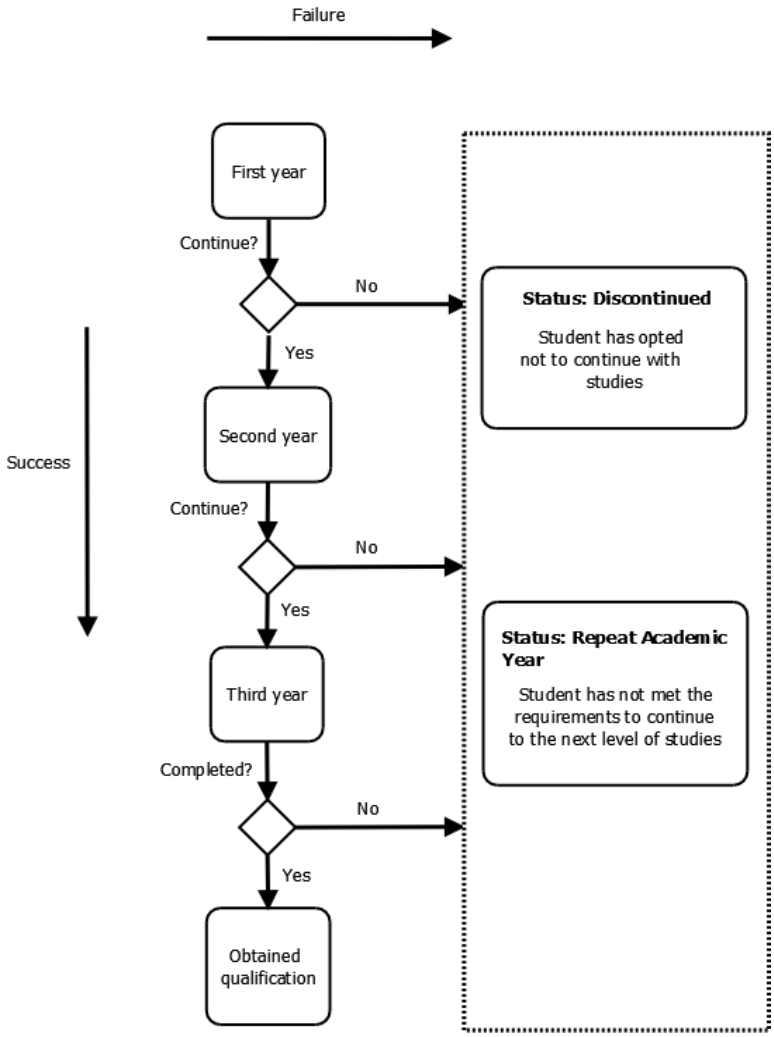

Fig. 2. Higher education trajectory of an academic qualification that spans three years depicting both the success and failure as it pertains to the higher education trajectory of students. A success is represented by the downward facing arrow and a failure is represented by the horizontal facing arrow.

When performing this research the following research questions are proposed: (RQ1) What procedure is best in representing a student's higher education trajectory when the data is of temporal nature? (RQ2) To what extent do the results from using a student's higher education trajectory become an effective alternative to determining an admission criterion; monitoring student progress post-admission and determining potential for graduation when compared to using the APS metric?

In order to answer the proposed research questions, this research experimented with secondary and higher education academic scores of students from computer science and mathematics majors. Various Bayesian models were learned using a popular score-based structure learning algorithm known as the hill-climbing structure learning algorithm. In addition, this research implemented the Bayesian Estimator to learn the parameters of the model. The resulting learned Bayesian models were compared with a model developed using expert knowledge (known as the ground truth model), using a probability distribution statistical distance measure in order to ascertain how well the Bayesian model was able to learn the relationships between course modules in a higher education trajectory that spans three years.

In addition, a graphical view of the computed statistical distance between the learned model with the ground truth model was presented to provide an illustration of the impact of the learned Bayesian model structures relative to the ground truth model for increasing sample sizes. Thus, the resulting learned models are used to determine student success trajectories in a higher education trajectory that spans three years.

Results from this research will contribute to further the application of Bayesian probabilistic modelling in the domain of predicting students' success upon analyses of their trajectory through higher education. In addition, this research will provide higher education administrators with a mechanism to employ guidance tools to assist students with monitoring their higher education progression, hence, optimising their chances of graduating on time.

The structure for the remainder of this research document will begin with covering related work that have made significant contributions in using machine learning to predict student performance. Section 3 will provide an outline of the procedure used in performing the experiment of this research. The findings from this research will be outlined and discussed in Section 4. Finally, Section 5 will conclude with a summary, contributions and recommendations for future work for this research.

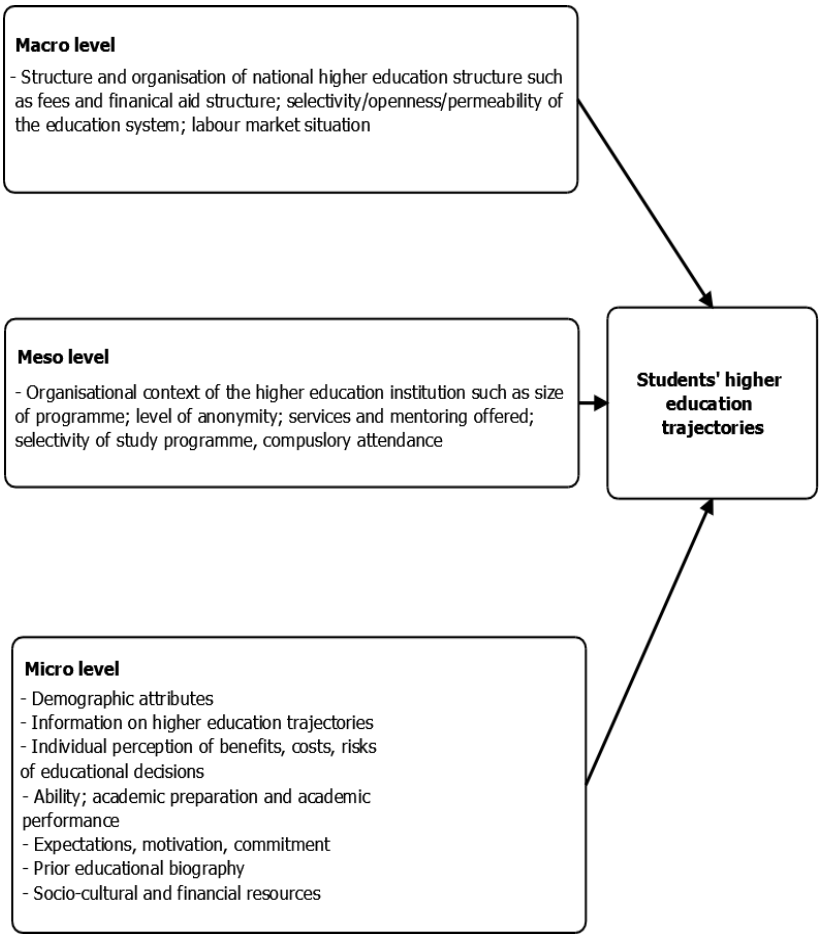

Fig. 3. Higher education trajectory predictors categorised into three different levels, namely: macro-level, meso-level and micro-level adapted by [4].

2. Related Work

The task of predicting student performance in an effort to identify students at risk of not obtaining their academic qualification on time has become an urgent requirement for educational institutions [9]. It is important that higher education institutions employ effective measures to improve student completion rates in response to the low on-time student graduation rate alluded to in the study by [2].

Results presented in Table 1. represent a summary of past studies that made useful contributions in employing Bayesian modelling to predict the performance of students. Furthermore, the results presented in the table make mention of the data that was used to perform the experiments; the features used in the development of the Bayesian model; the performance metrics used in evaluating the performance of the developed Bayesian models and the outcome from the performance metrics in the domain of predicting the success of students in school.

Furthermore, the features represented in Table 1. form the nodes in the construction of the respective Bayesian models. A relationship is established between the nodes during the learning Bayesian process illustrating the dependencies and independencies that exist between the nodes in the model. Furthermore, it is worth noting that the common features of interest amongst the observed studies in Table 1. are gender [6,10,7] and academic course grades [6,11,10,7,8] which form part of the micro-level predictors of predicting a student's higher education trajectory as illustrated in Fig.3..

This section will begin by exploring the impact of admission metrics that are currently used to admit students for higher education studies, in an effort to determine student success. This section will proceed to exploring studies that have made significant contributions in utilising machine learning models to aid in predicting the success of students. Finally, this section will conclude with studies that have utilised a probabilistic-graphical approach to determine the success of students and providing insights in the factors that affect the success of students.

Table 1. Bayesian learning approaches with their associated features; performance metrics and outcomes from the performance metrics determined from past studies when predicting student success.

Author	Data	Features (nodes)	Performance Metrics	Value
[6]	514 data records of high school students were considered.	Gender; group work attitude; interest for mathematics; achievement motivation; self-confidence; shyness; English performance and mathematics performance.	Accuracy	64%
[11]	129 data records of student performance and final grades for the academic course: Probability and Statistics.	Exam final grade; number of lectures attended in a certain week; number of practicums attended in a certain week; number of points awarded for achievement in the class in a certain week; quiz score; e-test score and an indicator of student persistent.	Weighted F-score Sensitivity	> 80% > 85%
[10]	Higher education students' data records from the faculty of science.	Grade; calendar year; year of study; gender; race; course; final grade outcome; progress outcome type description; country; citizenship; language; nationality; permit type description and institution description.	Accuracy	81.25%
[7]	A sample of 500 data records of higher education students.	Employment status; gender; semester; teacher and score.	Accuracy	66%
[8]	783 data records of first year higher education computer science students.	Province of secondary school; secondary school quintile; academic literacy and quantitative literacy results; financial aid status; science results; advanced mathematics attempt results; English results; work ethic; attitude and at-risk status	F1-Score MCC Sensitivity Specificity	84.14% 61.98% 66.67% 92.42%

2.1. Impact of Current Admission Techniques in Determining the Success of Students

It has been well-established that to admit prospective students for higher education studies, in response to the increase in access to higher education institutions [12], a viable admission technique is required to determine students that are most likely to complete their academic course successfully. This section will primarily focus on the contributions of admission metrics that are currently used to seek out candidates for higher education studies.

2.1.1. APS Metric as an Admission Indicator

In South African, the APS metric is one of the performance measures utilised by higher education institutions to seek out candidates for admission for higher education studies. According to a study by [3], almost 50% of students failed to complete the requirements of their academic course in the science stream.

Furthermore, an interesting result emanating from [3] showed that relying solely on the employment of the APS score to determine the success of a student in higher education did not yield the desired student success rate as compared to incorporating biographical and individual attributes to the predictive model. Table 2. illustrates this point with results of the predictive accuracy of the aforementioned models. The predictive accuracy becomes greater when incorporating both biographical and individual attributes as opposed to relying solely on the APS score and the combination of the APS score with the grades for English and Mathematics as predictors for determining the success of students in higher education [3].

Table 2. Predictive accuracies resulting from a study by [3] for determining learner vulnerability for different combinations of features as opposed to using the APS score solely.

Attributes	Machine learning model predictive accuracy		
	Multilayer perception	Logistic regression	Naive Bayes
APS score	30%	34%	28%
APS score; English grades; Mathematics grades	50%	44%	38%
APS score; English grades; Mathematics grades; Biographical information; Individual attributes	85%	83%	82%

2.1.2. Grade Point Average as an Admission Indicator

The Grade Point Average (GPA) metric is a common admission criteria administered by higher education institutions because of its perceived importance in measuring the academic outcome of students in higher education level [13]. A study conducted by [13] experimented with both science test results and English test results in addition to the secondary school GPA to predict the higher education outcome of students admitted to a medical programme. The study found that the science test results proved to be a good predictor of performance for medical students as compared to the other predictors considered [13].

2.1.3. Admission Techniques Summary

Results produced from studies in these sections provide an indication that solely depending on the secondary school GPA or APS score is not viable as the only indicator to admit students for higher education studies. It is important to introduce valid and reliable admission tools [13] in order to improve in the intake of prospective students for studying at higher education institutions.

2.2. Machine Learning Approach to Predicting Student Success

Results from Table 1. and Table 2. place significant emphasis on the predictive capabilities attached to using machine learning models to determine student success. In order to make use of the predictive capabilities of machine learning models, it is important to select the right features that have an effect on determining the success of students in higher education by employing feature selection techniques.

2.2.1. Feature Selection for Maximising the Predictive Accuracy

A popular feature selection technique utilised by past studies in order to enhance the predictive capabilities of a machine learning model is the information gain (entropy) method. Reference [14] attempted to use information gain (entropy) to determine the importance of each feature used in the study and their contribution in the event of determining the risk profile of students.

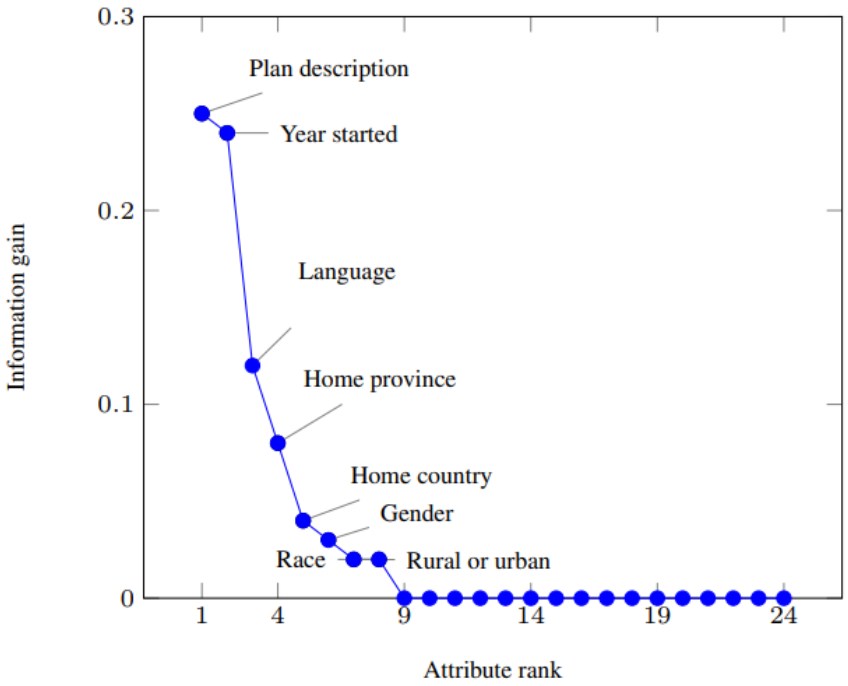

Fig. 4. Attribute ranking from the most important attributes to the least important attributes based on the computed information gain (entropy) when determining the risk profile of students adapted from [14].

Attributes with the highest information gain (entropy) are considered important attributes. In this study, the important features were both individual and biographical data as presented in the graph in Fig.4. [14]. By using the information gain (entropy) feature selection technique, the random forest model trained in the experiment by [3] produced the greatest accuracy of 85% which was significantly better than the accuracy produced from solely using the APS score as illustrated in Table 2.

Furthermore, there are numerous other studies that have utilised feature selection techniques to select the best features for their respective machine learning models. In particular, [15] experimented with a variety of feature selection techniques including: cforest; random forest; LMG; First; Last; Multivariate adaptive regression splines (MARS) and Boruta. Furthermore, [15] developed multiple multivariate linear regression models from the resulting top five attributes from each method respectively. It was found that the best multivariate linear regression model used in determining student performance resulted from applying the MARS method [15].

In addition to the computational aspect of selecting important features, another methodology applied in seeking out important features is to approach experts in the education field to assist in determining the most important factors that affect the success of students. One such study that applied this approach was [6]. The study mentioned above utilized expert opinions and literature reviews to select the top eight attributes for constructing a Bayesian model aimed at predicting performance in Mathematics. This effort resulted in a 64% accuracy rate [6].

2.2.2. Conceptual Framework

The results from [3] has shown that a student's APS score alone is not a viable admission metric to determine their success in higher education, hence, additional attributes are required to supplement the current admission criteria to determine the success of a student in higher education. The conceptual framework, depicted in Fig.5. proposed by [16], promotes a methodology of grouping student attributes into three groups, namely: background, individual and pre-college/school attributes. This framework has been successfully adopted by previous studies [17,3,14] to determine the success of students in their higher education studies.

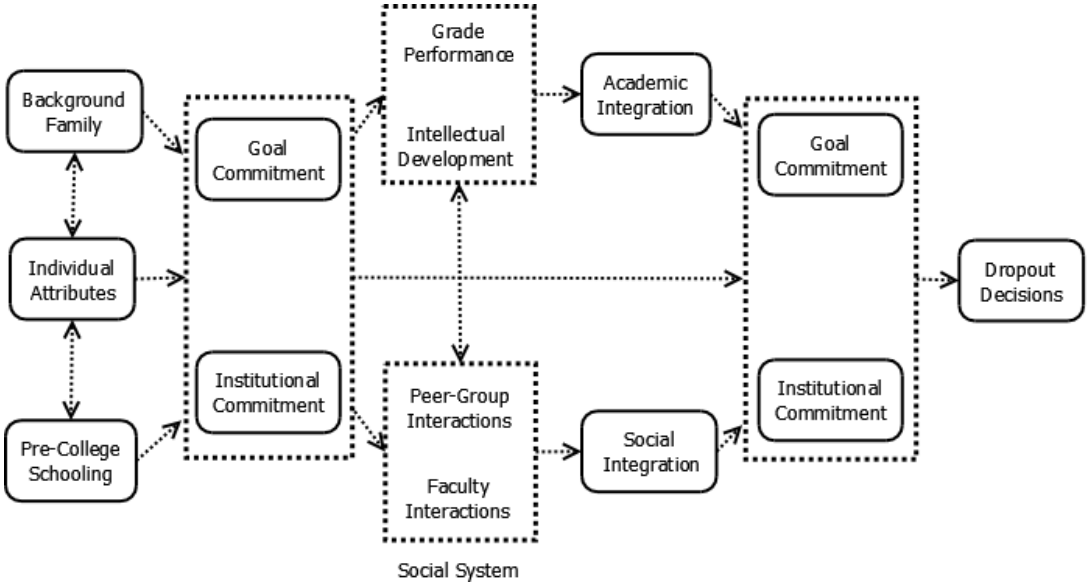

Fig. 5. Conceptual framework proposed by [16] presenting a system for using student and institutional attributes to determine the dropout status of students.

The results produced in Table 3. illustrate a grouping of attributes from past studies that experimented with machine learning models on a defined set of attributes to determine student success. The attributes from these studies are grouped into the three groups proposed in the conceptual framework by [16]. The introduction of other attributes into a predictive model, in addition to the APS score, has the potential of improving the predictive accuracies of the proposed machine learning model with the overall aim of determining student success in higher education.

Table 3. Grouping attributes used from past studies into background, individual and pre-college/schooling as per the conceptual framework defined by [16] to observe attribute group selection with model performance. Attributes shaded with blue for each associated study represents the attributes used to predict student performance in each study.

Author	Background	Individual	Pre-college/schooling	Best Model	Model Performance
[15]	✔	✔		Multivariate linear regression	▪ 4.368 (RMSE) ▪ 0.103 (R^2) ▪ 3.301 (MAE)
[18]	✔	✔	✔	Classification and Regression Tree	▪ 0.86 (Accuracy) ▪ 0.86 (F1-score) ▪ 0.97 (AUC)
[17]	✔	✔	✔	Naive Bayes	▪ 0.69 (Accuracy)
[19]			✔	Support Vector Classifier	▪ 0.88 (Accuracy)
[20]	✔	✔	✔	Multilayer Perceptron	▪ 0.974 (Accuracy) ▪ 0.974 (F1-score) ▪ 0.0190 (MAE) ▪ 0.0802 (RMAE) ▪ 0.0719 (RAE) ▪ 0.222 (RRSE)

Furthermore, it is evident from the resulting performance measures for each study that attribute selection plays a significant role in the predictive capabilities of a machine learning model. Thus, this provides an indication that the inclusion of more attributes (provided that these attributes make a significant contribution in predicting student performance) in a machine learning model increases the predictive capabilities of a machine learning model as opposed to depending on the APS score as the sole admission metric for determining student success.

The limitations of the non-probabilistic graphical models depicted in Table 3. lie in their inability to produce a graphical model. A graphical model would depict the dependent relationships that exist between the student results in a higher education trajectory that spans the length of an academic course. Thus, the graphical model and associated probability distributions associated with the nodes of the graphical model would be applied to infer the probability of success in a certain model given prior results within the trajectory.

For this reason, the results that are of particular interest to this research are those related to studies concerned in the implementation of probabilistic graphical approaches in order to determine the success of students in higher education as presented in Table 1. [6,11,10,7,8] .

2.3. Influence of Probabilistic-Graphical Models in Predicting Student Success

The probabilistic-graphical nature of the associated Bayesian models have contributed to understanding the probability of success or failure for students based on the relationships that are formed between students' attributes in an academic course using concepts in graph theory and probability theory.

This section will serve to unpack contributions made from applying probabilistic-graphical models in the domain of determining student success in higher education. Upon unpacking these contributions, this section will proceed to identifying further contributions to be made that will be attempted in this research.

2.3.1. Naive Bayes Model for Predicting Student Success

The naive Bayes model has proven itself as a simple but effective classification machine learning model when predicting student success. For example, [17] experimented with six machine learning models to develop a recommendation system for students, aiming to provide them with advice on their higher education trajectory. Based on the six machine learning models, the naive Bayes model resulted as the best model achieving the greatest accuracy of 69.18% [17].

The naive Bayes model may seem a viable machine learning model to determine student success from a probabilistic-graphical approach based on its simple yet effective nature. A probabilistic-graphical approach refers to using probability theory and graph theory to represent the uncertainty and complexity which one encounters in student data ensuring explainability of conditional dependencies.

Unfortunately, for the purposes of this research, the naive Bayes model is limited by its independence assumption.

The conditional independence assumption of a naive Bayes model suggests that all the attributes in a Bayesian model are conditionally independent of each other for a given class label for each given instance as illustrated in the diagram of a simple naive Bayes model in Fig.6. [3]. The diagram in Fig.6. shows that for a given class, nodes X_1 to X_n are conditionally independent. A Bayesian model that is not restricted by the independence assumption ensures that there is a deeper understanding of the relationship that exists between the students' attributes that may not have been apparent when implementing a naive Bayes model.

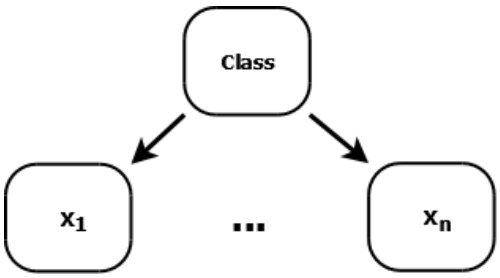

Fig. 6. A simple representation of a naive Bayes model where node features X1 to Xn is conditionally independent for a given class adapted by [3].

2.3.1.1 Bayesian Model Structure

The Bayesian model structure assists in visualising and identifying the relationships between the students' attributes in predicting student success. The selected attributes are represented by the nodes in the developed Bayesian model.

As opposed to placing reliance on expert knowledge to learn a Bayesian model structure, it is possible to learn the structure of a Bayesian model through employing one of a variety of structure learning algorithms. For instance, [7] made use of different algorithms including K2 and BDeu (Bayesian Dirichlet equivalent uniform) to construct a Bayesian model structure. An example of a Bayesian model structure resulting from [7] has been illustrated in Fig.7.

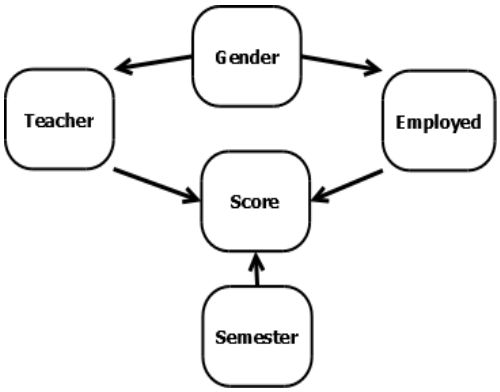

Fig. 7. The resulting Bayesian model structure used in predicting student scores adapted by [7].

Furthermore, it is possible to have some of the selected student attributes not present in the final Bayesian model structure. The student attributes not selected for the final model have no relationship with any of the other attributes. For instance, the Bayesian model structure developed by [10] experimented with fourteen biographical and enrollment student attributes and it was found that two of the attributes, permit type description (Permit) and institution description (SCHL) were presented with no relationships with the other attributes based on the illustrated results in Fig.8..

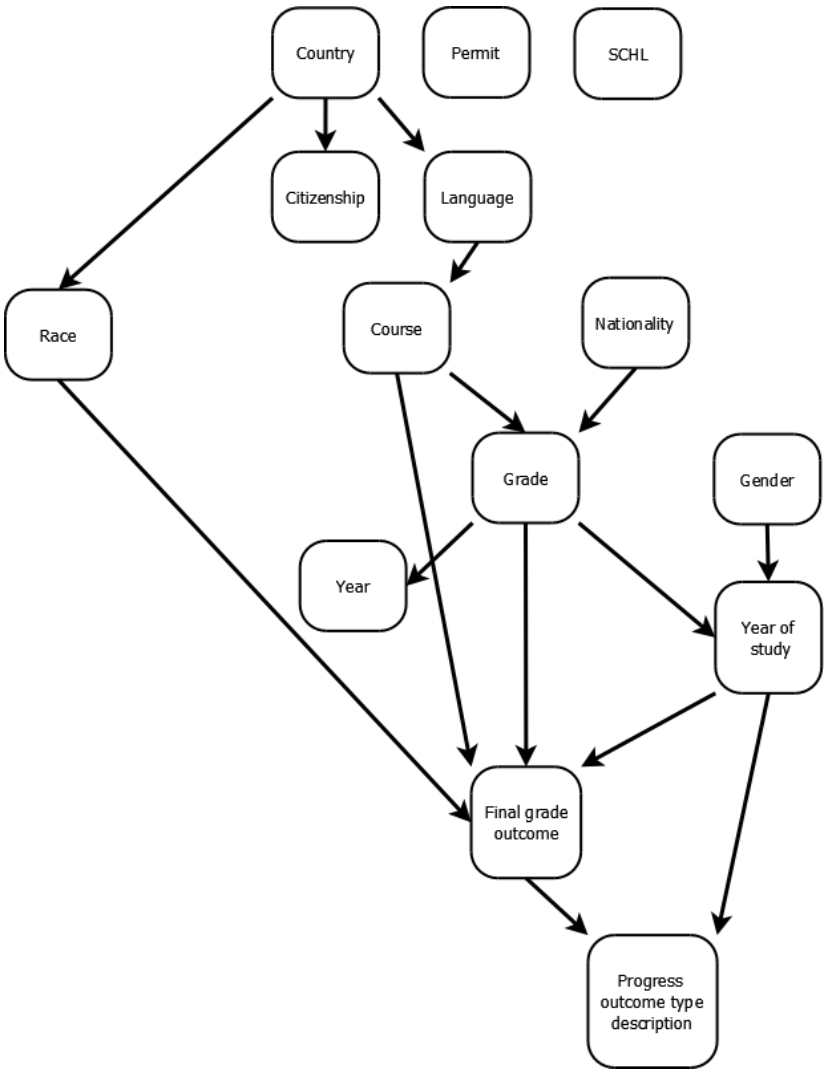

Fig. 8. The resulting Bayesian model structure used in predicting student scores adapted by [10].

2.3.1.2. Bayesian Model Parameter Estimation

The nodes representing each random variable (student attribute) of the Bayesian model have an associated continuous probability distribution presented in a continuous probability table (CPT). Reference [8] used the training data to compute prior probabilities for each random variable and thereafter the conditional and posterior probabilities were computed. In order to account for the immeasurable latent variables present in the construction of the Bayesian model, [8] opted to use the Expectation Maximization (EM) algorithm for the purpose of estimating the parameters of the developed Bayesian model.

There are a variety of other parameter estimation algorithms that are used for the purpose of computing the associated probability distribution for each random variable. Reference [10] employed the Bayesian parameter estimation method which functions by formulating the prior probability distribution on a random variable θ and uses the dataset to compute the posterior probability of the random variable θ.

2.3.1.3. Bayesian Model Inference

Bayesian inference allows for the possibility of employing a Bayesian model B for the purpose of computing the marginal probability $P(N=n)$, where n represents the sample size, for each node $v \in B$ and each possible instantiation v [21]. In the study by [10], a Junction Tree algorithm was used to infer students' grade based on their demographic attributes. [7] experimented with a number of inference algorithms including likelihood sampling; logic sampling and clustering to query the developed Bayesian model in order to determine a student's grade.

2.4. Future Work Based on Related Work

The resulting Bayesian model structures depict a graphical understanding of the relationships between the student attributes that lead to identifying at-risk students [11,8] and predict student success [10,7]. This may appear effective at face value, however, more work is required to further unpack results emanating from the application of Bayesian models and possibly improve on the predictive accuracy associated with determining a student's risk [8].

Taking into consideration results from this section and the need to further unpack these results, this research will attempt to explore the prediction of student success trajectories using a Bayesian approach in an academic course that spans three years. The design of the Bayesian model will be a dynamic model showing relationships between course modules for mathematics and computer science between first year and third year for a three year trajectory. In addition, certain elements from previous work including the choice of Bayesian estimator and score-based function will be incorporated in this research.

With that said, the next section will provide an in-depth methodology to conduct the experiment for this research including: the data construction; the development process of a probabilistic-graphical model; the usage of a statistical distance measure to compare the probability distributions of the probabilistic graphical models and conclude with the choice of inference algorithm to compute the posterior probability distributions.

3. Methodology

The work studied by past researchers, in utilising Bayesian modelling to determine student success, employ variables limited to only a single time period. This suggests that the studies do not model the students' higher education trajectory from their first year to their final year of studies. As such, these studies are limited in their capability to predict a successful higher education trajectory thereby improving the overall graduation throughput. The approach of this research is to employ a dynamic Bayesian model where the student results in mathematics and computer science are present in all the time slices of the model. As opposed to previous studies, the variables in this model are not static as they change for each time period. The non-static nature of these variables makes it possible to observe the change in results in these modules as students progress in their higher education trajectory that spans a period of three years with the goal of predicting a successful trajectory.

Furthermore, this research attempted to learn a variety of Bayesian models using the hill-climb structure learning algorithm and compare their learned probability distributions with the ground truth model using the Kullback-Leibler (KL) divergence metric. The significance of a methodology of this nature is to obtain a Bayesian model that is as close as possible to the ground truth model by measuring the statistical distance in their respective associated probability distributions. Thus, the obtained Bayesian model is used in determining student success trajectories in response to a call to further unpack results emanating from [8] by further observing higher education student trajectories.

Fig.9. is a high level illustration of the methodology used in this research. The methodology begins with a pre-processing of the data extracting the results associated with the computer science course. A ground truth model representing a student's higher education trajectory is developed using knowledge from educational experts. Thus, this research attempts to learn the higher education trajectory and evaluates the quality of learned trajectory relative to the trajectory represented by the ground truth model using the KL Divergence metric. Higher education administrators can use the learned trajectory based on inference computations for student admission; post admission monitoring and completion of a degree in order to improve completion rates and reduce dropout rates.

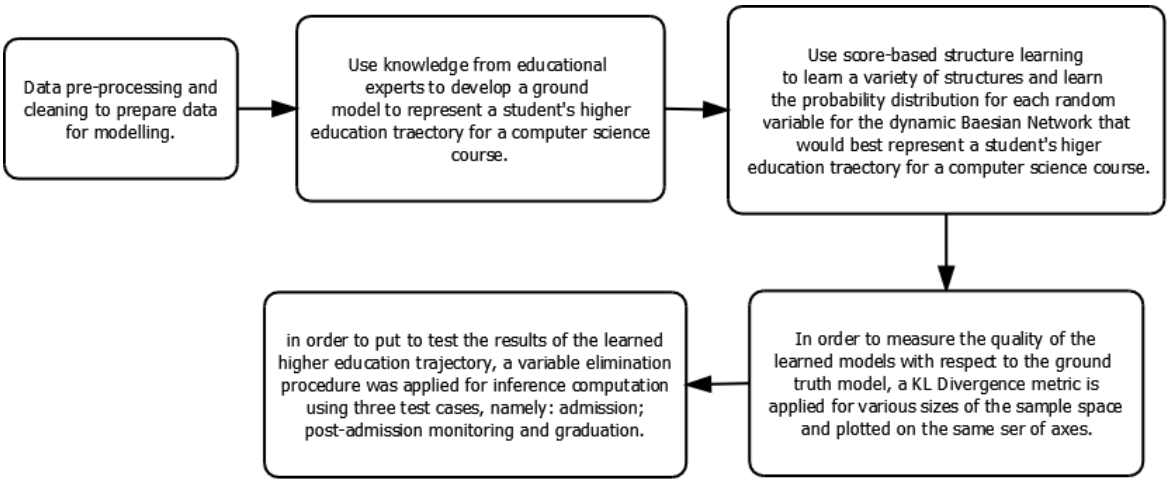

Fig. 9. The high level methodology for this research..

The outline of the methodology in the sections will proceed to further discuss data construction in order to get the data ready for the construction of the Bayesian model. Next, the development of the ground truth Bayesian model and learned models will be discussed along with algorithms used for structure learning; parameter estimation and inference. Finally, the probability distribution comparison between the ground truth model and learned models using the KL divergence metric will be discussed along with the observed effect of the computed average KL divergence with the increasing sample size of the dataset.

3.1. Data Construction

The data used for this research was synthetic data. Previous studies observed biographical and individual attributes in addition to the academic performance as per the conceptual framework proposed by [16]. The aim of this study was to predict the higher education trajectory of students purely from an academic performance point of view. As such, the attributes of importance for this research was the academic results from computer science and mathematics majors. The academic results are for Grade 12; first year; second year and third year. Thus, this highlights the student's academic higher education trajectory for a three year period.

In order to prepare the data for modelling, it was important to extract the results for subjects/modules contributing to the computer science for a 4 year period (Grade 12, First year, second year, third year) in order to build a dynamic Bayesian model. The extracted results would then constitute as random variables for the model. The dynamic Bayesian model would model the trajectory of results from Grade 12 to final year of university for the computer science degree by learning the associations between different time slices. Missing values would be handled by replacing the missing values (result) with the average for the respective subject/module. The results were then discretized based on the range of the results. The range [0 - 50) would represent a *fail*; [50 - 75) would represent a *pass* and [75 - 100] would represent a *pass with distinction*.

3.2. Probabilistic Graphical Model: Bayesian Model

According to [22], a probabilistic graphical model is defined as a multivariate statistical model born out of a marriage between graph theory and probability theory. There are instances in educational research such that there is uncertainty in the existence of associations between subjects/modules in a course and the ability to graphically represent the complexities associated with these associations. Hence, the graphical property associated with applying probabilistic graphical models is useful in representing uncertainty, whereas, the probabilistic property is useful in representing complexity [22].

A popular probabilistic graphical model applied by numerous studies (refer to Table 1..) is the Bayesian Network model. Based on the results from these studies, it is clear that the application of Bayesian Networks to predict a student's performance is just as effective as its black box model counterparts (refer to Tables 2. and 3.) based on the results of the accuracies produced in the model. A simple Bayesian model is characterized by nodes represented using random variables and edges representing the associations between the random variables. For the purposes of educational research, the nodes in the Bayesian model are represented by the student's marks whereas the edges are represented by the conditional dependencies between the student's marks in a higher education trajectory that spans three years. The nodes in the Bayesian model are represented by the student's marks with each node represented by a conditional probability distribution learned from data. The edges are represented by the conditional dependencies between the student's marks in a higher education trajectory that spans three years. According to [25], two random variables X and Y are said to be conditionally independent of each other if following the introduction of a known set Z, then knowledge of X provides no information about Y as illustrated in Fig.10..

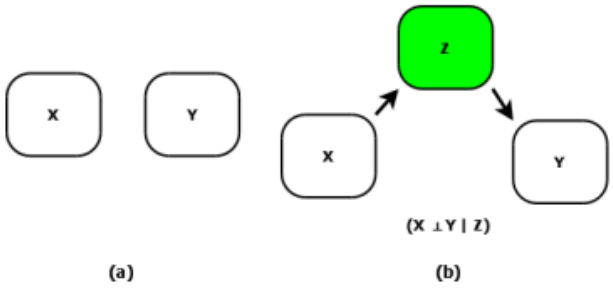

Fig. 10. An illustration depicting the conditional independence of two variables in a Bayesian model. (a) Represents the two conditional independent random variables *X* and *Y*. (b) Represents the addition of a set *Z* to confirm the conditional independence assumption ($X \perp Y \mid Z$) of the two random variables *X* and *Y*.

The significance of applying a Bayesian Network as compared to its black box model (these are models such that there is no explainability associated with the result) counterparts lies in its ability to model the uncertainty and complexities that you would find in student data. The uncertainty is using probability theory to do inference by computing the probability of producing a result (R) of fail pass or pass with distinction for a module given results from other modules or the status of other student attributes known as the evidence variables (E).

Furthermore, Bayesian Networks have the ability to incorporate prior knowledge into the model. The inference probability computation would be represented as $P(R|E_1, E_2, \dots, E_n)$ for n given evidence variables. The complexity in student data is represented using graph theory to graphically represent the learned associations between the results from each module or student attribute.

The graphical property of a Bayesian model lies in its directed acyclic nature. The directed acyclic nature of the Bayesian model graph implies that cycles in the graph are absent. The edges are represented by directional arrows as illustrated in the example of the resulting Bayesian model from a study by [6] in Fig.11..

Due to the acyclic nature of a Bayesian model, its associated graph is referred to as a directed acyclic graph (DAG). In addition to the directed acyclic nature of Bayesian models, there are also the conditional probability tables [21] associated with each node in the Bayesian model. The development of a Bayesian model is based on the Markov condition which states that every variable in a Bayesian model is independent of its non-descendants non-parents given its parents and leads to its joint probability distribution, given by the equation $P(X_1,\dots, X_n) = \prod_i P(X_i|parent(X_i))$, where $X_i, i = \{1,2\dots,n\}$, represents the independent variable in a Bayesian model [23].

Constructing a single time-slice Bayesian Network, as outlined by [7], involves defining the random variables. The random variables defined in this research are the results from the computer science course. The conditional dependencies between these random variables are learned and the associated conditional probability distribution for each random variable is computed.

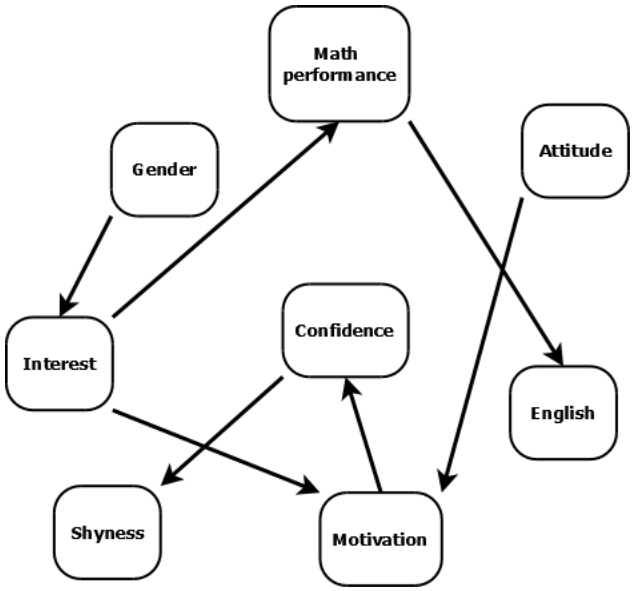

Fig. 11. Example of a Bayesian model resulting from a study by [6].

The random variables that make up the Bayesian models in this section are static. This implies that they do not change. The Bayesian models in this section serve as an entry point to the specialised form of a Bayesian model which is the dynamic Bayesian model. The next section extends from static random variables and discusses random variables

that change over time which are essential for the purposes of this research given that the concept of higher education trajectories are introduced. The variables chosen for this research are time-dependent, hence, the decision for the application of a dynamic Bayesian model to model the trajectory of a student in a three year degree.

3.3. Dynamic Bayesian Model

A dynamic Bayesian model describes a system that evolves with time [24]. As such, the static nature of the standard Bayesian model, discussed in Section 3.2, becomes a dynamic model as it now models motive forces [24].

The Markov assumption and time invariant assumption (defined at-length in the study by [25]) compactly represent a trajectory over time and are used to unroll the transition model (contained in the two-time slice Bayesian model along with the initial distribution) into a dynamic Bayesian model [25].

The graphical representation of a dynamic Bayesian model is made up of variables occurring at multiple time slices with edges emanating within each time slice and different time slices. According to [25], intra-time-slice edges are edges that occur within each time slice whereas inter-time-slice edges are edges that occur between time slices. Furthermore, edges that occur between the same variables in different time slices are known as persistent edges and the variables are known as persistent variables [25]. An example of a dynamic Bayesian model, adapted from [25], unrolled over three time slices is illustrated in Fig.12..

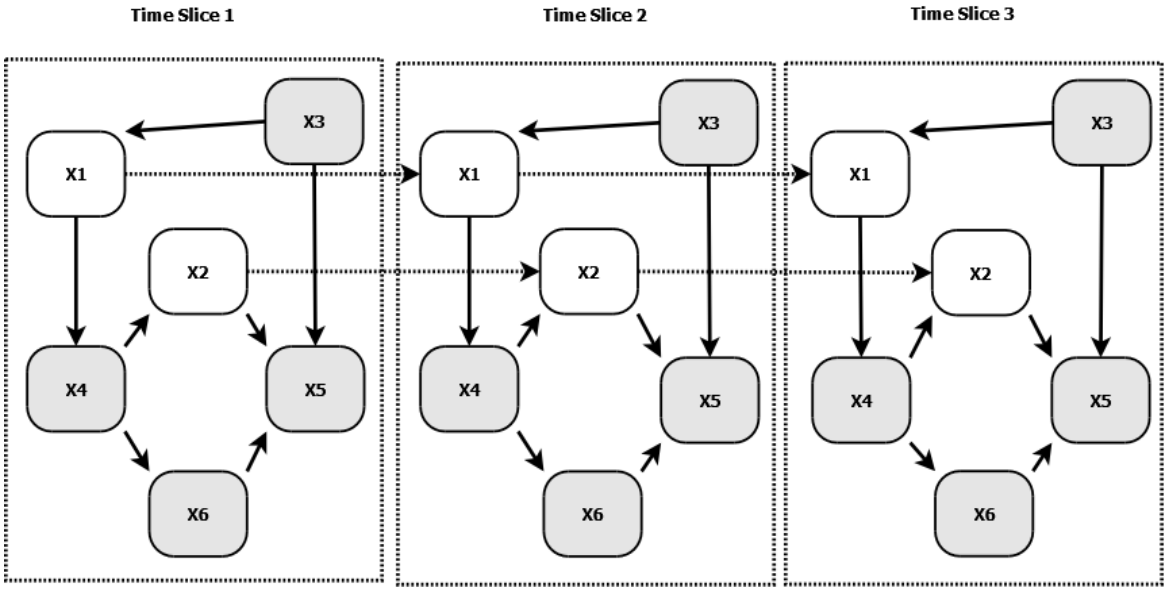

Fig. 12. An example of a dynamic Bayesian model unrolled over three time slices. The shaded variables are observable whereas the unshaded variables are persistent variables as they are connected to each other in adjacent time slices using the dotted edges. This illustration has been adapted by [25].

Furthermore, according to [24], the probability distribution function associated with a dynamic Bayesian model system is represented by the equation $P(X,Y) = \prod_{t=1} P(x_i|x_{t-1}) \prod_{t=1} P(y_i|x_t)P(x_0)$. The description of the parameters of the dynamic Bayesian equation are as follows: T represents the time boundary; $X= (x_0,..,x_{T-1})$ represents the sequence of T hidden-state variables; $Y = (y_0,..,y_{T-1})$ represents the sequence of T observable variables; $P(x_i|x_{t-1})$ represents the state transition probability distribution function; $P(y_i|x_t)$ represents the observation probability distribution function and $P(x_0)$ represents the initial state distribution.

3.3.1. Construction of a Ground Truth Bayesian Probabilistic Model

The development of the ground truth model for this research was through expert knowledge and from literature. The developed ground truth model, illustrated in Fig.13., is represented as a dynamic Bayesian network with four time slices. Time slice 0 represented the student's Grade 12 results whereas time slices 1, 2 and 3 represented the student's results for first, second and third year respectively in a higher education trajectory that spans three years.

The course codes *MA*, *CS* and *AG* represent the final mathematics, computer science and aggregate score results respectively for a student in a higher education trajectory that spans three years. There are relationships formed between the academic scores within the time slices and between adjacent time slices. The dotted lines are indicative of the inter-time slice persistent edges. The computer science and mathematics scores are referred to as persistent variables whereas the final aggregate score is the observable variable.

This research has considered only results for mathematics; computer science and the aggregate results for the purpose of modelling the change in results for the three year degree. Based on the availability of data, it is possible to consider other student related factors, for example, students' mode of learning for the duration of their degree. However, due to limitations in the dataset, this research is limited to only the students' results.

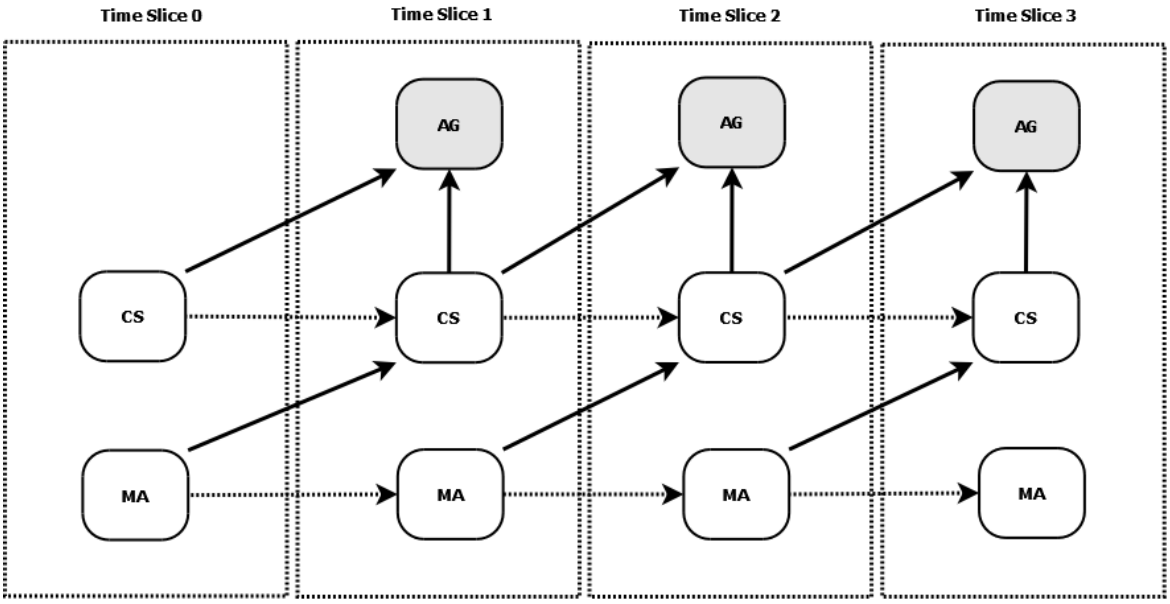

Fig. 13. The ground truth Bayesian model structure for this research with four time slices. Time slice zero represents the grade 12 results while time slices 1, 2 and 3 represents the student's higher education results for first, second and third year respectively. The computer science and mathematics scores (unshaded) represent the persistent variables due to the dotted persistent edges whereas the final aggregate score (shaded) represents the observable variable.

Upon forming relationships between adjacent time slices, it is important to note that the variables (states) represented by the student's mathematics; computer science and aggregate scores, in the dynamic Bayesian model in Fig.13. satisfy the Markovian condition such that the state at time slice *t* depends only on the states at its immediate time slice and current time slice [24]. In addition to the valid relationships between nodes within each time slice, the valid relationships formed between nodes in different time slices are those that exist in adjacent time slices and must be from a lower time slice *l* to a higher time slice *l+1* adjacent to the previous time slice *l* as illustrated in the example in Fig.14..

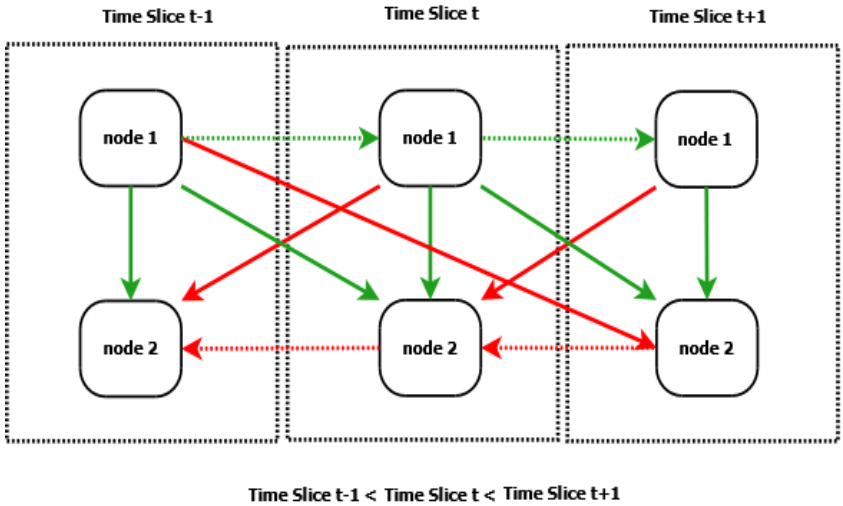

Fig. 14. Example of an illustration of the Markovian condition for a dynamic Bayesian model from time slice t-1 to time slice t+1. The green directional arrows represent the valid relationships between the nodes whereas the red directional arrows represent the invalid directional arrows as it pertains to the satisfaction of the Markovian condition.

3.3.2. Construction of a Learned Bayesian Probabilistic Model

This stage of the research methodology attempted to learn the structures of a variety of Bayesian models using the popular score-based hill-climbing structure learning algorithm. According to [26], learning a Bayesian model is an NP-hard problem which justifies the usage of a heuristic search algorithm such as the hill-climbing structure learning algorithm.

The score-based hill-climbing structure learning algorithm was implemented using the following score functions: BDeu score; BIC score; BDs score; K2 score and AIC score in order to learn the structure of the Bayesian model. The importance of the score function lies in its ability to determine how well the data fits to the model [22].

In addition, score-based algorithms put forward a criterion in which the structure of the Bayesian model can be evaluated on the proposed dataset [22]. As such, all possible Bayesian model structures are searched and scored and the Bayesian model structure with the highest possible score is chosen as the best structure for the Bayesian model [22].

3.4. Bayesian Parameter Learning using Bayesian Estimation

The goal of estimation is to use sample data to estimate parameters of a model. The capability associated with Bayesian estimation is its ability to incorporate prior knowledge with new data to estimate parameters of the model regardless of the complexity of the model [27].

This research utilised the Bayesian estimation estimator in an attempt to learn the parameters of the Bayesian model. In order to observe the impact of learning the associated probability distributions for the Bayesian model with different sample sizes of the dataset, the parameters of the Bayesian model are learned for different sizes of the sample dataset and compared with the ground truth model using a statistical distance metric.

Bayesian estimation considers the parameters (θ) as random variables as opposed to considering them as unknown constants [10]. A prior probability distribution $p(\theta)$ is formulated and is used to compute the posterior probability $p(\theta|D)$ from the dataset D [10]. According to [25], the values of the parameter θ updates as more data is obtained over time.

3.5. Kullback-Leibler (KL) Divergence Metric for Comparing Probability Distributions

Upon constructing the ground truth model and learning the Bayesian model from data, a KL divergence is computed to compare the probability distributions associated with the ground truth model with that of the learned models. The KL divergence computes the error made in learning the Bayesian model by measuring the difference in the probability distributions between the learned model and the ground truth model [28].

KL Divergence is a useful metric utilised in this research for measuring the quality of the learned model with respect to the ground truth model by computing the difference in probability distributions associated with each random variable using the equation provided in (1). This research learned a dynamic Bayesian model by computing the probability distributions associated with each random variable represented in a continuous probability table (CPT). Consequently, the KL Divergence was applied to learn the quality of the learned dynamic Bayesian model by measuring the difference in the probability distribution of the learned model with respect to the probability distribution of the ground truth model by using the probability distributions in the CPT associated with each random variable. This will be able to ascertain how well the dynamic Bayesian model was learned with respect to the ground truth model. The average KL Divergence is computed for different sizes of the sample space for a variety of structure scores (BDs, BDeu, BIC and K2) and these computed KL Divergence values are plotted on the same set of axes.

The computation of the KL divergence (denoted as $KL_{G,L}$), adapted from [28], between the ground truth model (G) and the learned model (L), defined on the same set of variables $X = (x_1, x_2 \ldots, x_n)$, is presented in (1).

$$\begin{aligned} KL_{G,L} &= \sum_x p^G(x) \log\left(\frac{p^G(x)}{p^L(x)}\right) \\ &= \sum_x p^G(x) \left[\log\left(p^G(x)\right) - log\left(p^L(x)\right)\right] \\ &= \sum_x p^G(x) \log\left(p^G(x)\right) - \sum_x p^G(x) log\left(p^L(x)\right) \end{aligned} \tag{1}$$

The purpose of the KL divergence statistical measure is to determine the statistical difference between the probability distribution associated with the ground truth model (G) which is represented by $p^G(x)$ and the probability distribution associated with the learned model, (L) which is represented by $p^L(x)$.

3.6. Inference for Bayesian Models

The inference process for Bayesian models involves computing the probability of the variable of interest based on the condition of a given evidence or set of fixed variables [29]. The aim of inference is to compute the posterior probability *P(X|E)* of the variable of interest (state variable) *X* given the observable evidence *E* as illustrated in Bayes theorem in equation (2).

The probability of the evidence variable *E* given the state variable *X* represented as *P(E|X)* is known as the likelihood. The probability of the state variable *X*, represented as *P(X)*, is known as the prior probability whereas the probability of the evidence *E*, represented as *P(E)*, is known as the marginal probability. As such, the combination of these parameters give rise to the equation in Bayes theorem, (2), which is the basis for the computation of the posterior probability distribution *P(X|E)* in inference calculations.

$$P(X|E) = \frac{P(E|X)P(X)}{P(E)} \tag{2}$$

Several inference algorithms exist to compute the posterior probability of the state variable given the evidence. Inference algorithms are categorised into two categories, namely: exact inference algorithms and approximate inference algorithms. The examples of the two categories of inference algorithms are represented in the illustration in Fig.15. adapted by [30]. The inference algorithms colour-coded in green represent the exact inference algorithms whereas the approximate inference algorithms are colour-coded in yellow.

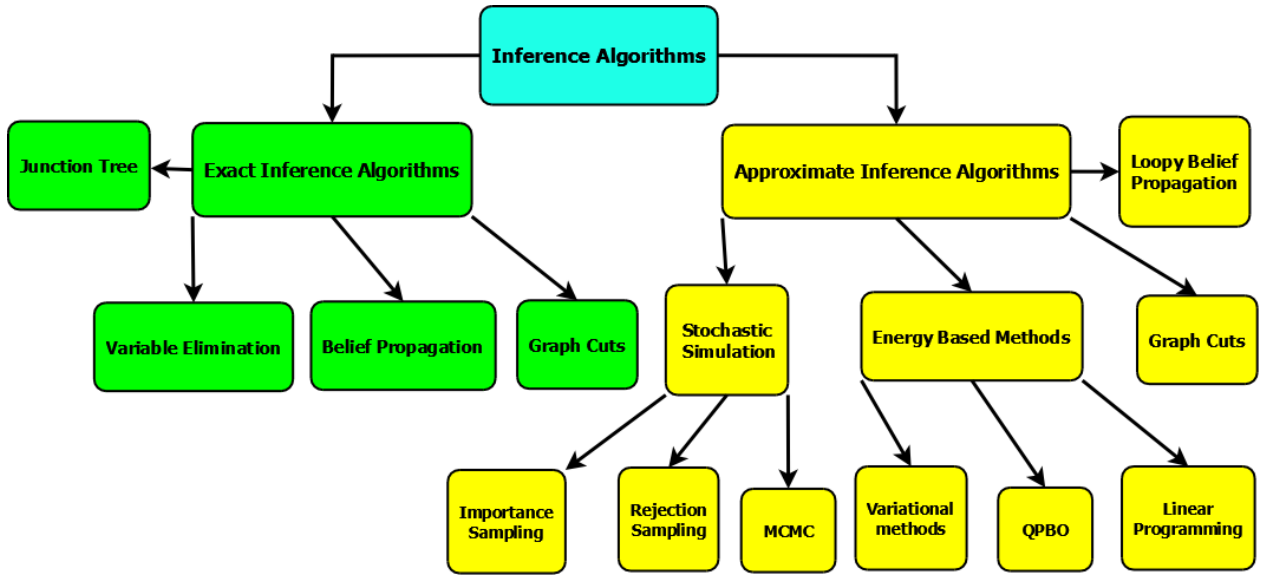

Fig. 15. Examples of inference algorithms categorised into exact inference algorithms and approximate inference algorithms adapted by [30].

This research implemented an exact inference algorithm known as variable elimination to compute the posterior distributions. The inference results aimed to justify the reasoning for Bayesian probabilistic modelling to predict student performance as opposed to the APS metric currently employed to gauge the performance of a student.

The reason for implementing the variable elimination inference algorithm is due to its simplistic nature [31]. It is considered a fundamental probabilistic inference algorithm when employed on Bayesian models [32]. The inference computations performed in this research are fairly simple test cases to illustrate the advantage of probabilistic-graphical modelling to determine student success as opposed to placing reliance solely on the APS metric. As such, for the purposes of this research, variable elimination is the sufficient choice for inference computation.

3.6.1. Exact Inference

For many real world applications, exact inference is known to be tractable and is limited by its performance which is deemed exponential in the worst case [29]. According to [30], for arbitrary joint distributions defined over a larger group of random variables, exact inference is known to be intractable.

This category of inference takes an evidence e and initial variable v and uses them to compute the likelihood of e ($P(e)$) and the marginal distribution ($P(V = v)$) or posterior distribution ($P(V = v \mid e)$) of v [21]. The purpose behind the implementation of exact inference algorithms is to manipulate the joint distribution in order to reduce the overall computational complexity [30]. The exact inference algorithm discussed and employed in this research is the variable elimination inference algorithm.

3.6.1.1. Variable Elimination

According to [31], variable elimination is an exact reasoning algorithm with the ability to solve complex reasoning associated with Bayesian models. The complexity associated with using the variable elimination methodology lies in the order of its elimination with the detection of the optimal elimination order classified as an NP-hard problem [31].

The algorithm for executing the variable elimination operation, adapted from [33], has been presented in the pseudocode in Algorithm 1. The operations of importance in the variable elimination procedure are the factor multiplication and factor marginalization in which the detailed definitions can be found in the study by [33].

According to Algorithm 1, the procedure for the variable elimination algorithm requires a set of conditional probability distributions (CPDs) that are used to construct the factors (also referred to as potentials) $\mathbf{\Phi} = \{\Phi_1,...,\Phi_N\}$. The operations that are possible on the factors are marginalisation (summation) of a variable and factor multiplication. Other requirements to implement the variable elimination algorithm include the query variables X to compute the posterior distribution; the observable variables y and the fixed elimination variables Z.

Algorithm 1. Variable Elimination Algorithm Adapted By [33]

Require a set $\prod$ of CPDs associated with the Bayesian model, the query variables X in which to compute the posterior distribution and the evidence variables y which are considered the observed variables

Using the CPDs contained in the set $\prod$, the factors $\mathbf{\Phi} = \{\Phi_1,..., \Phi_N\}$ are constructed
The variable set is partitioned into query X, evidence y and elimination Z
for all $i = 1,...,N$ **do**
    Replace Φ_i with Φ_i $[Y=y]$
end for
From the variables contained in Z, the ordering $Z_1,...,Z_k$ are fixed
for all $i = 1,...,k$ **do**
    $\Phi' \leftarrow \{\Phi \in \mathbf{\Phi} : Z_i \in Scope\ (\Phi)\}$
    $\Phi'' \leftarrow \Phi \setminus \Phi'$
    $\Psi \leftarrow \prod_{\Phi \in \Phi'} \Phi$ - This is the factor multiplication procedure
    $\rho \leftarrow \int_{Zi} \Psi(\cdot) dZ_i$ - This is the factor marginalization procedure
    $\Phi \leftarrow \Phi'' \cup \{\rho\}$
end for
return Φ

The fixed elimination variables stored in Z are ordered. According to [31], in terms of the computational cost of the elimination method, the cost for elimination is considered to be $\delta = \Sigma_{i=1}\delta(Z_i)$ where $\delta(Z_i) = |Z| \prod_{i=1}|X_i|$. Hence, $\delta(Z_i)$ is referred to as the elimination cost of the variable Z [31]. The computation of the query involves summing out the variables in Z one at time [32]. Upon completing the factor multiplication and marginalisation of the variable Z procedures, the algorithm returns the new factors $\mathbf{\Phi}$ with the variables from Z removed from the initial factors.

According to [29], the computation of variable elimination is contributed by eliminating all variables stored in Z one at a time after an ordering and operation has been performed over the potentials. The potentials refer to the local distributions of the variable and the parents of the variable [29]. An example illustration of the variable elimination process, illustrating the elimination of two variables, has been presented in Fig.16. adapted by [29].

The illustration in Fig.16. is a representation of the graphical procedure of eliminating the variables coherence and difficulty. It is clear from the diagram that the order of elimination is the variable coherence first followed by the variable difficulty. The variable elimination procedure terminates when there are no longer variables to eliminate and returns the variables with the absence of the eliminated variables.

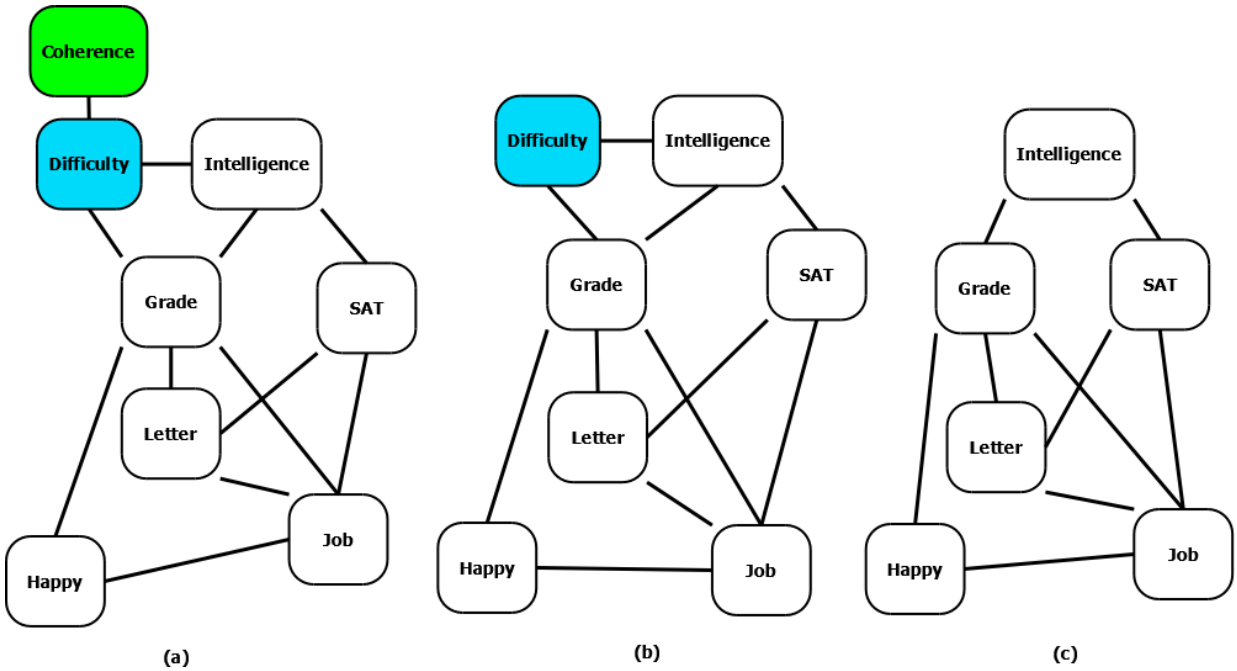

Fig. 16. An example illustration of the variable elimination process adapted by [29] where (a) represents the variables *coherence* and *difficulty* in the list of variables before they are eliminated; (b) represents the elimination of the variable *coherence* and (c) represents the elimination of the variable *difficulty*.

The next section will present and discuss the results from experiments conducted from this methodology including the KL divergence results; variable elimination inference results; results as they pertain to answering the proposed research questions and the limitations of this research.

4. Results and Discussion

The expectation from implementing the outlined methodology is to approach the prediction of student success trajectories from a probabilistic-graphical point of view by employing Bayesian probabilistic modelling. The hill-climbing structure learning algorithm is employed to learn the structures of a variety of Bayesian models and compare the probability distribution associated with the learned models and the ground truth model using the KL divergence statistical measure. The final model selected is as close to the ground truth model based on their associated probability distributions in order to use it to determine student success.

The ground truth model illustrated by Fig.13. represented the true higher education trajectory developed based on expert knowledge. The structure learning algorithms employed to learn the learned models to compare with the ground truth model included: the K2 score-based model; the BDeu score-based model; the BDs score-based model and the BIC score-based model. The Bayesian estimation estimator was employed to learn the probability distributions associated with the nodes of the learned Bayesian models and produce a continuous probability table (CPT).

The results are indeed theoretical when using synthetic data; however as a future work this research can be built upon with real student data and tested for different courses other than computer science. Furthermore, these results are centered on showcasing a methodology for building a dynamic Bayesian model and using it to associate explainability with student success. The success is interpreted as graphical depiction of a student's higher education trajectory with associated dependencies between the adjacent time slices and being able to compute probabilistic inferences using a temporal point of view.

This section will explore the outcomes emanating from the results in comparison with results from previous studies by implementing the outlined methodology of this research. Thus, these results formulate the answers for the proposed research questions. In addition, this section will discuss the limitations associated with implementing this research.

4.1. Results in Comparison with Previous Studies

Results from previous studies that made use of machine learning were formulated to find good predictors of performance and evaluated the strength of these predictors relative to the applied machine learning model using a performance metric [15,19,18,20]. For example, [15] explored background and individual features to predict student performance using a multivariate linear regression model and used the RMSE, R^2 AND MAE performance metrics to evaluate the strength of the predictors relative to the using multivariate linear regression to predict student performance.

Previous studies that introduced probabilistic graphical models not only predicted student performance but also introduced the concept of explainability associated with using a Bayesian Network [6, 11, 10, 7, 8]. With the results from the Bayesian Network, one can introduce prior knowledge into the model when performing prediction computations.

In order to continue this progression, this research introduced dynamic Bayesian Networks which have the capability of modeling student data that have a temporal nature which is something you cannot do with a static Bayesian Network. The dynamic Bayesian Network made it possible to represent a student's higher education trajectory (RQ1) and use the model to perform inference computations based on prior knowledge and learned probability distributions associated with each random variable (RQ2).

This research learned a variety of score-based structures (K2, Bdeu, BIC, BDs) with the associated CPTs for each random variable in the structure. This was done to learn the best structure to depict the ground truth model which was modeled from knowledge from educational experts. The variety of learned structures was compared with the ground truth model by measuring the learned probability distributions associated with each random variable using the KL Divergence metric and plotted on the same set of axes as observed in the results in Fig.17..

4.2. KL Divergence Probability Distribution Comparison

The aim of employing a statistical distance measure is to be able to ascertain how well the learned models learned by comparing these models with a true model. The statistical distance measure employed in this research was the KL divergence metric. An average KL divergence was computed between the developed Bayesian ground truth model and the various learned models, constructed using the hill-climbing structure learning algorithm, to compare the probability distribution associated with the academic modules for the students when constructing the Bayesian probabilistic model.

The results presented in Table 4. illustrate that the average KL divergence decreases as the sample size increases. This suggests that the probability distributions computed during the learning process for the various learned models closely resembles the ground truth model for large sizes of the sample dataset. Furthermore, based on the results presented in Table 4., the BDs score-based Bayesian model resulted as the Bayesian model that closely resembled the ground truth model by a small margin when compared to the other learned Bayesian models.

The resulting graph is illustrated in Fig.17. graphically represents the impact of the sample size of the dataset with the computed KL divergence metric relative to the ground truth model represented by the blue curve. For relatively

large sample sizes, the average KL divergence reaches its minimum for all Bayesian models considered including: the BDs score-based model; the BDeu score-based model; the BIC score-based model and the K2 score-based model.

Table 4. Average KL divergence comparing the probability distributions associated with the ground truth model and various learned models for varying sample sizes of the dataset.

Sample size	Bayesian models			
	BDs score model}	BDeu score model	BIC score model	K2 score model
100	0.33597	0.30192	0.31801	0.28345
1000	0.16072	0.14104	0.16332	0.15022
2000	0.12809	0.12760	0.13123	0.12502
3000	0.12277	0.10872	0.11997	0.10979
4000	0.09476	0.10293	0.10861	0.10519
5000	0.09191	0.10134	0.10338	0.09668
6000	0.08964	0.09275	0.10101	0.09757
7000	0.08769	0.09726	0.09630	0.09026
8000	0.08893	0.09490	0.09480	0.09472
9000	0.08629	0.09244	0.09243	0.09228

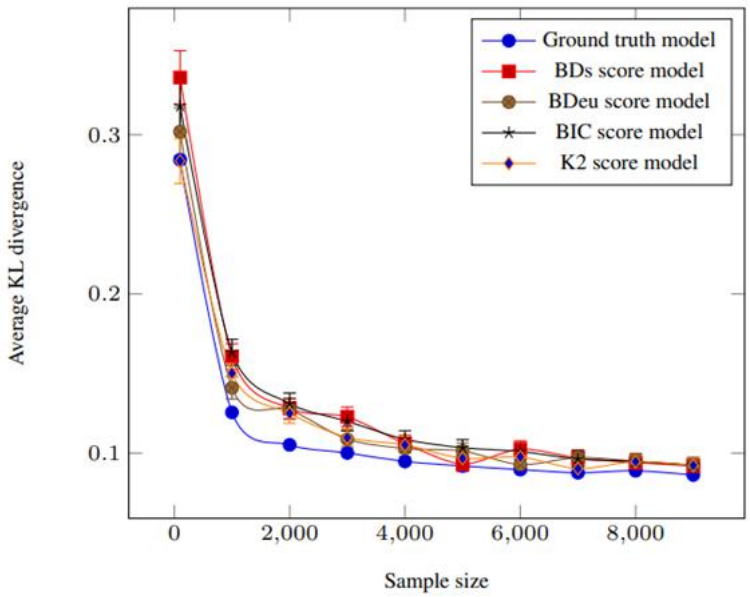

Fig. 17. Average KL divergence comparing the probability distributions associated with the ground truth model and various learned models with increasing sample size with the addition of error bars (5% error).

The learned probability distributions associated with each of the aforementioned Bayesian models constructed using the hill-climbing structure learning algorithm has shown to be closer to the ground truth model considering the KL divergence plot in Fig.17. Thus, these results provide an indication that it is plausible to consider the hill-climbing structure learning algorithm to learn a Bayesian model to predict student success trajectories.

In addition to representing the probability distribution difference between the various learned models and the ground truth model, Fig.17. has incorporated error bars, using 5% error, to each of the points in the curves of the various learned Bayesian models. The addition of error bars represent the variability of the computed average KL divergence and are an indication of the error made in computing the KL divergence for increasing sample sizes of the dataset for each learned Bayesian model.

An interesting observation in Fig.17. is that as the sample size increases the lengths of the error bars get smaller for all Bayesian models considered. This is an indication that the learned model gets better at learning and resembling the true model as the sample size increases.

Previous literature studied in Section 2 that modelled student results, particularly literature concerned with the employment of probabilistic graphical models, utilised either the naive Bayes model or the static Bayesian model. The random variables of the naive Bayes model operate under the independence assumption whereas the random variables of the static Bayesian model do not change over time. This research took it a step further by exploring random variables that change over time, hence, the employment of a dynamic Bayesian model.

The advantage of implementing the dynamic Bayesian model over the static Bayesian model is the ability to model the trajectory of students' results over the duration of the degree. In addition, the inclusion of the KL divergence computation provides clarity on the learned model that best resembles the ground truth model when comparing their respective probability distributions. In the case of this research, based on the results presented in Table 4. and Fig.17., the BDs score-based Bayesian model was found to resemble the ground truth model in comparison to the other learned models.

Given that the BDs score-based Bayesian model closely resembled the ground truth model based on the computed average KL divergence values in Table 4., the following subsection will attempt to employ the BDs score-based Bayesian model in order to perform inference computations using the variable elimination inference algorithm. This is done so as to illustrate the impact of probabilistic-graphical modelling on higher education trajectories on student admission; student monitoring and to gauge the graduation status of students as opposed to relying solely on the APS metric.

4.3. Variable Elimination Inference Results

This research employed the variable elimination inference algorithm to illustrate the significance of probabilistic-graphical modelling on determining student success as opposed to the APS metric. This is performed by computing the posterior probability distribution of a test case given the evidence variables. The test cases in the following subsections depict student admission; student monitoring post-admission and gauge graduation status using a probabilistic-graphical approach by employing the learned BDs score-based Bayesian model.

4.3.1. Test Case 1: Student Admission

Consider a student with Grade 12 results as depicted in Table 5. The results depicted are for mathematics and computer studies. The aim was to use these given evidence variables to compute the posterior probability distribution for first year success in higher education studies in order to determine the admission status of the student. The variable to query is the final aggregate score for the first year.

Table 5. Student results used as evidence variables in the variable elimination inference algorithm for test case 1.

Subject	Outcome
Grade 12 mathematics	Pass
Grade 12 computer studies	Fail

The result from the variable elimination inference algorithm performed on the student's results in Table 5. when using the BDs Bayesian model is depicted in the form of a table (Table 6.) and a graph (Fig.18.). The result suggests that a student with the results (evidence variables) as depicted in Table 5. will most likely pass the first year as the majority class distribution is the *pass* class with a probability of 0.8301. The next class is the *fail* class with a probability of 0.1374 and the lowest class is the *pass with distinction* class with a probability of 0.0325. These results suggest that the student is less likely to pass first year with a distinction.

Table 6. Results from the variable elimination inference algorithm depicted using the BDs score-based Bayesian model for test case 1 depicting the probability distribution for a student's performance in first year.

First year final aggregate score	Φ (First year final aggregate score)
Fail	0.1374
Pass	0.8301
Pass with distinction	0.0325

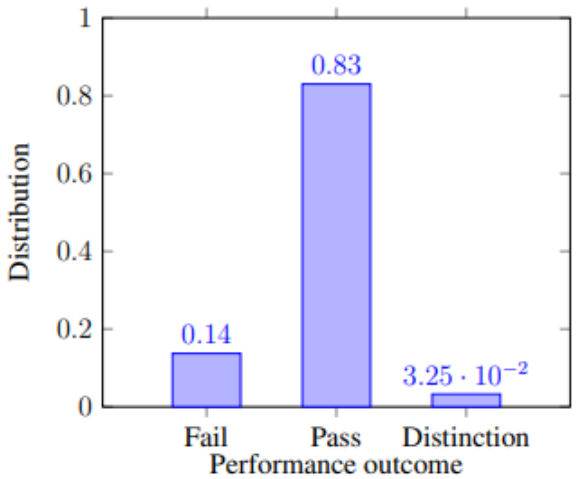

Fig. 18. Results from the variable elimination inference algorithm using the BDs score-based Bayesian model for test case 1 depicting the probability distribution for a student's performance in first year.

4.3.2. Test Case 2: Post-Admission Monitoring

Consider a student with results as depicted in Table 7.. The results depicted are for Grade 12 mathematics and computer studies; first year mathematics and computer science and first year aggregate score. The aim was to use these given evidence variables to compute the posterior probability distribution for second year success in higher education studies as a means to further monitor a students' academic progress in their higher education trajectory. The variable to query is the final aggregate score for second year.

The result from the variable elimination inference algorithm when using the BDs Bayesian model is depicted in the form of a table (Table 8.) and a graph (Fig.19.). The result suggests that a student with the results (evidence variables) as depicted in Table 7. will most likely pass the second year as the majority class distribution is the *pass* class with a probability of 0.7885. The next class is the *fail* class with a probability of 0.1936 and the lowest class is the *pass with distinction* class with a probability of 0.0179. These results suggest that the student is less likely to pass second year with a distinction, however, the student is likely to pass the second year of study.

Table 7. Student results used as evidence variables in the variable elimination inference algorithm for test case 2.

Subject	Outcome
Grade 12 mathematics	Pass
Grade 12 computer studies	Fail
First year mathematics	Pass
First year computer science	Pass with distinction
First year aggregate score	Pass

Table 8. Results from the variable elimination inference algorithm depicted using the BDs score-based Bayesian model for test case 2 depicting the probability distribution for a student's performance in second year.

Second year final aggregate score	Φ (Second year final aggregate score)
Fail	0.1936
Pass	0.7885
Pass with distinction	0.0179

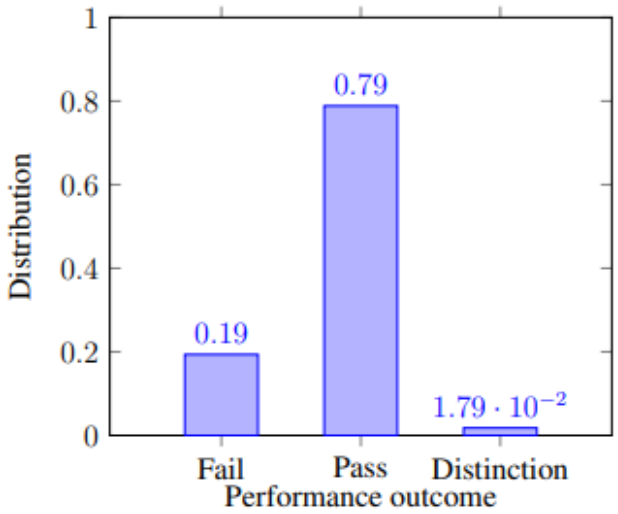

Fig. 19. Results from the variable elimination inference algorithm using the BDs score-based Bayesian model for test case 2 depicting the probability distribution for a student's performance in second year.

4.3.3. Test Case 3: Monitor Student Completion Status

Consider a student with results as depicted in Table 9. The results depicted are for Grade 12 mathematics and computer studies; first year mathematics and computer science; first year aggregate score; second year mathematics and computer science and second year aggregate score. The aim was to use these given evidence variables to compute the posterior probability distribution for third year success in higher education studies as a means to gauge the completion status of the student. The variable to query is the final aggregate score for third year.

Table 9. Student results used as evidence variables in the variable elimination inference algorithm for test case 3.

Subject	Outcome
Grade 12 mathematics	Pass
Grade 12 computer studies	Fail
First year mathematics	Pass
First year computer science	Pass with distinction
First year aggregate score	Pass
Second year mathematics	Pass
Second year computer science	Pass
Second year aggregate score	Pass

The result from the variable elimination inference algorithm when using the BDs Bayesian model is depicted in the form of a table (Table 10.) and a graph (Fig.20.). The results suggest that a student with the results (evidence variables) as depicted in Table 9. will most likely pass the third year as the majority class distribution is the *pass* class with a probability of 0.7632. The next class is the *fail* class with a probability of 0.2021 and the lowest class is the *pass with distinction* class with a probability of 0.0348. These results suggest that the student is less likely to pass third year with a distinction, however, the student is likely to pass the third year of study.

Table 10. Results from the variable elimination inference algorithm depicted using the BDs score-based Bayesian model for test case 3 depicting the probability distribution for a student's performance in third year.

Third year final aggregate score	Φ (Third year final aggregate score)
Fail	0.2021
Pass	0.7632
Pass with distinction	0.0348

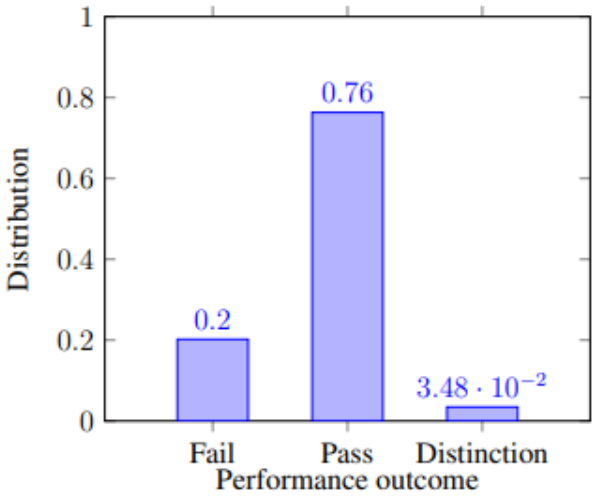

Fig. 20. The results from the variable elimination inference algorithm are depicted using the BDs score-based Bayesian model for test case 3 depicting the probability distribution for a student's performance in third year.

4.3.4. Test Case Results Summary

The aim of the three test cases depicted in this section was to illustrate the significant impact of probabilistic-graphical models in determining student success. The three test cases depicted were student admission; post-admission monitoring and gauge student completion. The outcome from these test cases were the probabilistic distributions from the three classes of performance outcome including: fail, pass and pass with distinction.

The results further emphasize the importance of viewing a student's performance in a simplified probabilistic-graphical point of view. Higher education institutions can employ this probabilistic-graphical approach to improve student admission and monitoring post-admission thereby improving the overall graduation throughput.

The APS system still remains a popular choice as an admission criterion because its computation takes into account the student's final results for Grade 12. Each subject is associated with a score based on the recorded result for that subject. The scores are added up for each subject and the resultant sum is the APS score for that student. The APS metric is useful in the admission process as it remains one of the potential predictors of performance to gauge the potential of a student's success in first year.

However, solely relying on the APS system raises some concerns on its predictive nature. The APS metric becomes effective when used in conjunction with other predictors of student performance as illustrated in the results by [3] in Table 2. The results from this research further improves on this as higher education administrators will not only

concern themselves with the admission process but also be involved in the student's trajectory to graduation which is the main contribution from this research from a higher education success point of view.

The next subsection will make use of the results produced from this research to answer the research questions proposed for this research.

4.4. Results as they Pertain to the Research Questions

4.4.1. Research Question 1

The graph illustrated in Fig.17. represented the computed average KL divergence between a variety of learned models and the ground truth model for varying sample sizes of the dataset. A methodology of this nature, which resulted in the results produced in this research, constitutes a viable course of action as it suggests that the learned probability distribution associated with the random variables of the developed Bayesian models can be compared with that of the ground truth model using a statistical distance metric in order to ascertain how well the Bayesian models were able to learn relative to the ground truth model. Thus, this action is able to represent a student's higher education trajectory.

Given this result, it is plausible to make use of these learned models to represent a student's higher education trajectory and use inference algorithms, such as the variable elimination inference algorithm employed in this research, to compute the associated performance distribution of the student.

4.4.2. Research Question 2

The advantage of using a Bayesian probabilistic model is that it provides an effective simplified probabilistic view of a student's outcome as opposed to using the APS metric. The graphical nature of the developed Bayesian model depicts the dependencies and independencies that exist between a student's results in a higher education trajectory that spans three years.

Furthermore, the results from the variable elimination inference algorithm performed to compute the posterior distribution based on the given evidence highlights a probabilistic view of determining student success as opposed to depending solely on the APS metric. The examples performed in the test cases in the variable elimination inference section illustrated the performance distribution for fail; pass and pass with distinction.

This suggests that a probabilistic view of students' academic progress has the potential of assisting higher education institution administrators gauge their potential for admittance to an academic programme based on their predicted academic trajectory and monitor their academic progress post admission stage on-route to graduation. This is in line with the aim of improving the overall on-time graduating rate.

4.5. Discussion Summary

The purpose of this research was to approach determining the success of student trajectories from a probabilistic-graphical point of view by employing Bayesian probabilistic modelling. The rationale for this was as a result of growing concerns over high dropout rates and low graduation throughputs in higher education institutions [1]. In addition, the purpose was to find an alternative to the APS metric currently utilised by South African higher education institutions as a student admission requirement.

The features used in constructing the Bayesian model results emanated from computer science and mathematics majors. The ground truth model, depicted in Fig.13. was developed from expert knowledge and used as a base model to compare against when learning a Bayesian model. The ground truth model developed was a dynamic model with four time slices. Time slice 0 represented the Grade 12 results of a student whereas time slices 1, 2 and 3 represented the higher education journey of a student in first year; second year and third year in a higher education trajectory that spanned a trajectory of a period of three years.

This research explored Bayesian models implemented using various score functions of the popular hill-climbing structure learning algorithm. The score functions included: BDs score; BDeu score; BIC score and K2 score. In addition, the parameters of the learned models were computed using the Bayesian Estimation estimator for different sizes of the sample dataset.

The results from this research presented the computed average KL divergence between the probability distributions of the various learned Bayesian models with the ground truth model for increasing sample size. The results of the computed average KL divergence were presented in Table 4.. In terms of a graphical representation, the observed effect in Fig.17. displayed an exponential decrease in the average KL divergence with increasing sample size for each learned Bayesian model relative to the ground truth model. A representation of this nature suggests that for large sample datasets, the learned models get better at learning and closely resemble the ground truth model.

Furthermore, this research implemented an exact inference algorithm known as variable elimination. The purpose of implementing the inference algorithm was to compute the posterior distribution of a set query based on given evidence with the outcome of the algorithm presenting a probabilistic view of a student's aggregate score distribution.

The test cases presented in this research was to further emphasize the significance of the probabilistic-graphical approach to predicting student success as opposed to relying solely on the APS metric. The test cases included: student admission; post-admission monitoring and gauge student completion.

The outcome from the KL divergence results are indicative of how well the Bayesian models learned when compared with the ground truth model by comparing their associated probability distributions. The outcome from the inference algorithm is indicative of an alternative to viewing the higher education trajectory of a student and using it to compute the probability distribution of a student's final results as opposed to relying solely on the APS metric.

The significance of implementing a Bayesian Network is the explainability factor that is associated with it. The explainability element provides higher education administrators with a graphical representation of a student's higher education trajectory with associations between courses within each year and adjacent years. The results from this show that a student's higher education trajectory can be represented graphically from first year to final year and a structure can be learned to model this trajectory by using prior knowledge. Higher education administrators can use this trajectory to compute the probability that a student would do well in certain courses within a degree by using inference computations.

The results produced in the inference computations in section 4.3 using variable elimination provide an effective indication as to how higher education administrators can further interpret results from the learned model by using evidence (represented as known results from previous modules) to compute the probability of success for a certain module. Higher education administrators can use results from these probability inference computations to refine/update admission criteria for courses in the admission process (results from test case 1). Next, higher education administrators can use these results to monitor a student's progress post admission and offer advice for module selection to maximize the probability of success for the selected module based on past results (results from test case 2). Finally, higher education administrators can use these results to predict the probability of passing all final year modules so as to graduate (results from test case 3). It is this complexity and uncertainty that comes with student data that makes Bayesian Networks the ideal choice to model a student's higher education trajectory that can prove to be helpful for higher education administrators.

4.6. Research Limitations

The results for this research are based solely on the contents of the synthetic dataset generated for this research. Thus, it has not been tested with a dataset that represents the real world student environment which would be beneficial in confirming these results. Experimenting on real world data is beneficial in gaining more insights into a student's higher education trajectory in a practical environment so as to put in place interventions in higher education institutions to maximise the probability of graduating on time.

Furthermore, previous studies that predicted student success using black box models were limited to finding predictors of student performance but lacked in computing the conditional dependencies between the predictors and having a graphical depiction of the associations between predictors because of the complexity associated with the process [15,18,19,20].

In order to incorporate a probabilistic graphical model, there existed studies that used a Naive Bayes model to predict student performance [17]. However, Naive Bayes model is limited by its independence assumption. Hence, it made sense to implement a popular probabilistic graphical model known as a Bayesian Network to account for the complexity and uncertainty in computing the conditional dependencies between predictors of student performance and computing the associated probability distributions for each predictor to be used in inference computations. However, the single time-slice Bayesian Network is limited to variables that are static and do not change over time (there is no temporal nature associated with the random variables).

Thus, this research implemented a dynamic Bayesian Network to account for the temporal nature associated with the random variables of concern. The result of this allowed for modelling the trajectory of students' results depicting conditional dependencies within each time slice and within adjacent time slices using a graphical representation. It is worth noting that the implementation of a dynamic Bayesian Network becomes effective if the random variables being modeled change over time which is the case in this research.

5. Conclusion

The main takeaway from this research is that the development of a dynamic Bayesian model provides a clear picture of a student's higher education trajectory for a three year degree. In comparison, static Bayesian models would only view a student's performance at a single point in time. In addition to the APS metric (currently used to admit students for higher education studies), higher education administrators would benefit from a complete view of students' projected results for the degree of their choice in order to make the final admission decision.

The significance of employing a probabilistic-graphical model lies not only on its predictive capabilities, but also on the graphical nature that is prominent in gaining information and understanding associations that exist between the student attributes in predicting the success of students. As such, this research provides a simplified probabilistic-graphical view of predicting student success trajectories using Bayesian probabilistic modelling with the intention of improving the overall on-time graduation throughput.

There are opportunities for researchers to test the results from this research by using real world data. Testing the results of this research using real world data would provide more insights in determining student success in a practical environment.

In addition, this research made use of score-based structure learning algorithms to learn the Bayesian Network and compared the learned networks with the ground truth model. Other classes of structure learning algorithms can be implemented such as constraint-based structure learning algorithms which include: the PC algorithm; Grow-Shrink and Inter-IAMB as well as hybrid algorithms. It would be worth investigating how well these classes of structure learning algorithms (constraint-based and hybrid) compare to the ground truth model and that of score-based algorithms using the KL Divergence metric in predicting a student's higher education trajectory. Similar to score-based algorithms, the procedure would involve learning the associations and conditional probabilities associated with each random variable therefore producing a CPT.

Furthermore, this research can be extended to investigating the curriculum mapping problem. This problem involves incorporating all influences of curriculum design (macro, meso and micro influences) in designing a curriculum that would ensure the success of both students and institutions of learning. With student trajectories forming a part of curriculum mapping, it would be interesting to develop a model that would represent and map all the influences of curriculum design into a single structure and use it to generate curriculums for educational institutions. In terms of modelling, a multilevel model would be a great place to start when investigating the curriculum mapping problem given that the influences are from different levels.

References

[1] T. Omphile and M. Lindikaya, "Estimating South African higher education productivity and its determinants using fare-primont index: Are historically disadvantaged universities catching up?," Research in Higher Education,2022, vol. 64, pp. 1–22, doi: 10.1007/s11162-022-09699-3.

[2] R. Maluleke, "Education Series Volume V Higher education and skills in South Africa. Technical report," Statistics South Africa, 2017.

[3] R. Ajoodha, "Identifying academically vulnerable learners in first-year science programmes at a South African higher-education institution", South African Computer Journal, South Africa, 2022, vol. 34, pp. 120–148, doi: https://doi.org/10.18489/sacj.v34i2.832.

[4] C. Haas and A. Hadjar, "Students' trajectories through higher education: a review of quantitative research," Higher Education, 2020, vol. 79, doi: 10.1007/s10734-019-00458-5.

[5] H. Roslan and C. Jen Chen, "Educational data mining for student performance prediction: A systematic literature review (2015-2021)," International Journal of Emerging Technologies in Learning (iJET), 2022, vol. 17, pp. 147-179, doi: 10.3991/ijet.v17i05.27685.

[6] R. Bekele and W. Menzel, "A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students," IASTED International Conference on Artificial Intelligence and Applications, part of the 23rd Multi-Conference on Applied Informatics, Innsbruck, Austria, 2005, pp. 189-194.

[7] R. Torabi, P. Moradi and A. Khanteymoori, "Predict student scores using Bayesian networks," Procedia - Social and Behavioral Sciences, 2012, vol. 46, pp. 4476-4480, doi: 10.1016/j.sbspro.2012.06.280.

[8] Z. Nudelman, D. Moodley and S. Berman, "Using Bayesian networks and Machine learning to predict computer science success," 47th Annual Conference of the Southern African Computer Lecturers' Association, SACLA 2018, Gordon's Bay, South Africa, June 18–20, 2018, Revised Selected Papers, Springer CCIS 963, 2019, pp. 207-222, doi: 10.1007/978-3-030-05813-5_14.

[9] L. Mahmoud Abu Zohair, "Prediction of Student's performance by modelling small dataset size," International Journal of Educational Technology in Higher Education, 2019, vol. 16, pp. 18, doi: 10.1186/s41239-019-0160-3.

[10] M. Mathye and R. Ajoodha "A Bayesian approach to predicting student performance using biographical and enrollment observations," unpublished, University of the Witwatersrand, South Africa, 2018.

[11] T. Kustitskaya, A. Kytmanov and M. Noskov, "Student-at-risk detection by current learning performance indicators using Bayesian networks," 2020, doi: 10.48550/arXiv.2004.09774.

[12] M. Cross and C. Carpentier, "'New Students; In South African higher education: Institutional Culture, student performance and the challenge of democratisation," Perspectives in Education, 2009, vol. 27, pp. 6-18.

[13] A. Almarabheh, M. Shehata, A. Ismaeel, H. Atwa and A. Jaradat, "Predictive validity of admission criteria in predicting academic performance of medical students: A retrospective cohort study," Frontiers in Medicine, 2022, doi: 10.3389/fmed.2022.971926.

[14] N. Mngadi, R. Ajoodha and A. Jadhav, "A theoretical model to predict undergraduate learner attrition using background, individual, and schooling attributes," unpublished, University of the Witwatersrand, South Africa, 2020.

[15] Q. El Aissaoui, Y. El Alami El Madani, L. Oughdir, A. Dakkak and Y. El Allioui (2020), "A Multiple Linear Regression-Based Approach to Predict Student Performance," Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) , 2020, pp. 9-23, doi: 10.1007/978-3-030-36653-7_2.

[16] V. Tinto, "Drop-outs from higher education: a theoretical synthesis of recent research," Review of Educational Research, 1975, vol. 45, pp. 89-125, doi: 10.2307/1170024.

[17] T. Abed, R. Ajoodha and A. Jadhav, (2020) "A prediction model to improve student placement at a South African higher education institution,"2020 International SAUPEC/RobMech/PRASA Conference, 2020, pp. 1-6, doi: 10.1109/SAUPEC/RobMech/PRASA48453.2020.9041147.

[18] E. Buraimoh, R. Ajoodha and K. Padayachee, "Application of Machine Learning Techniques to the Prediction of Student Success," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-6, doi: 10.1109/IEMTRONICS52119.2021.9422545.

[19] K. Al Mayahi and A. Mahmood, "Machine learning based predicting student academic success," 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, 2020, pp. 264-268, doi: 10.1109/ICUMT51630.2020.9222435.
[20] V. Nedeva and T. Pehlivanova, "Students' performance analyses using machine learning algorithms in WEKA," IOP Conference Series: Materials Science and Engineering, 2021, vol. 1031, pp. 012061, doi:10.1088/1757-899X/1031/1/012061.
[21] S. Rama, K. Jaladi and D. Devarapalli, "Bayesian network and variable elimination algorithm for reasoning under uncertainty," IJCSIT, 2012, vol. 3, pp. 5227-5230.
[22] Y. Zhou, "Structure learning of probabilistic graphical models: A comprehensive survey," ArXiv, 2011, vol. abs/1111.6925.
[23] F. Cozman 2000, "Generalizing variable elimination in Bayesian networks," Workshop on Probabilistic Reasoning in Artificial Intelligence, 2020.
[24] V. Mihajlovic and M. Petkovic, "Dynamic Bayesian networks: A state of the art," Europhysics Letters (epl), 2001.
[25] R. Ajoodha and B. Rosman, "Influence modelling and learning between dynamic Bayesian networks using score-based structure learning", PhD thesis, The University of the Witwatersrand, South Africa, 2018.
[26] J. Gámez, J. Mateo and J. Puerta, "Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood," Data Mining and Knowledge Discovery, 2011, vol. 22, pp. 106-148, doi: 10.1007/s10618-010-0178-6.
[27] M. Zyphur and F. Oswald, "Bayesian estimation and inference: A user's guide," Journal of Management, 2013, vol. 41, pp. 390-420,doi: 10.1177/0149206313501200.
[28] S. Moral, A. Cano and M. Gómez-Olmedo, "Computation of Kullback– Leibler divergence in Bayesian networks," Entropy, 2021, vol. 23, pp. 1122, doi: 10.3390/e23091122.
[29] M. Tubia, C. Bielza and P. Larranaga, "Facilitating the inference interpretation in Bayesian networks," unpublished, Universidad Politecnica de Madrid, 2023.
[30] A. Grim and P. Felzenszwalb, "Message passing dynamics of belief propagation algorithms," PhD thesis, Brown University, 2022.
[31] D. Ren, J. Guo and X. Hao, "Bayesian network variable elimination method optimal elimination order construction," ITM Web Conf, 2022, vol. 45, pp. 01012, doi: https://doi.org/10.1051/itmconf/20224501012.
[32] C. Aslay, M. Ciaperoni, A. Gionis and M. Mathioudakis, "Query the model: precomputations for efficient inference with Bayesian networks," ArXiv, 2019, vol. abs/1904.00079.
[33] I. Simonsson, "Exact inference in Bayesian networks and applications in forensic statistics," PhD thesis, Chalmers University of Technology and University of Gothenburg, 2018.

Authors' Profiles

Thabo Ramaano received his BSc (Honours) degree in Big Data Analytics and MSc in Computer Science from the University of the Witwatersrand, Johannesburg, South Africa. He is currently a student at the University of the Witwatersrand, Johannesburg, South Africa. His research interests include the application of machine learning in understanding the factors that influence student success and applying probabilistic graphical approaches.

Ashwini Jadhav is a Senior Lecturer in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa. She is passionate about, and has extensive experience in the field of student academic success and the factors that affect their successful trajectories. Her research interests include modelling of environmental factors for sustainability, in line with the SDGs like Quality education, Climate action, clean water resources and Zero hunger.

Ritesh Ajoodha received the Ph.D. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa, in 2018. He is currently with the School of Computer Science and Applied Mathematics, University of the Witwatersrand. His research interests include the application of machine learning for the synthesis and understanding of art and culture, the modelling of environmental factors for sustainability of key resources necessarily for life, diagnosing academic and social issues in order to promote student success, and the probabilistic modelling and learning of influence between processes for knowledge discovery and density estimation.

How to cite this paper: Thabo Ramaano, Ashwini Jadhav, Ritesh Ajoodha, "Developing a Bayesian Network Model to Predict Students' Performance Based on the Analysis of their Higher Education Trajectory", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.1, pp. 21-47, 2026. DOI:10.5815/ijmecs.2026.01.02