International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 7, pp. 1937-1955
doi: 10.25147/ijcsr.2017.001.1.130
https://stepacademic.net

Short Paper*

Multivariate Predictive Modelling of Mathematics Semestral Grade via Bayesian Networks Machine Learning Algorithm

Norie Neil C. Acosta
College of Science and Computer Studies
De La Salle University – Dasmariñas, Cavite, Philippines
ann1577@dlsud.edu.ph
(corresponding author)

Christian Kalki M. Lamadrid
College of Science and Computer Studies
De La Salle University – Dasmariñas, Cavite, Philippines
lcm1810@dlsud.edu.ph

Rolando B. Barrameda
College of Science and Computer Studies
De La Salle University – Dasmariñas, Cavite, Philippines
rbbarrameda@dlsud.edu.ph

Date received: January 29, 2023
Date received in revised form: February 16, 2023
Date accepted: February 18, 2023

Recommended citation:

Acosta, N.N., Lamadrid, C.K., & Barrameda, R. (2023). Multivariate Predictive Modelling of Mathematics Semestral Grade via Bayesian Networks Machine Learning Algorithm. *International Journal of Computing Sciences Research*, 7, 1937-1955. https://doi.org/10.25147/ijcsr.2017.001.1.130

Special Issue on International Research Conference on Computer Engineering and Technology Education 2023 (IRCCETE 2023). Guest Associate Editors:* **Dr. Nelson C. Rodelas, PCpE (*Computer Engineering Department, College of Engineering, University of the East-Caloocan City; nelson.rodelas@ue.edu.ph*) and **Engr. Ana Antoniette C. Illahi**, PCpE (*Asst. Professor & Vice Chair, Department of Computer and Electronics Engineering, Gokongwei College of Engineering, De La Salle University, Taft Ave., Manila; ana.illahi@dlsu.edu.ph*).

Abstract

Purpose – This study pertains to the novel use of Bayesian Networks to elucidate the interplay between multivariate factors of demographics, personality types, mathematics anxiety, and study habits in predicting the mathematics grades of college students. The research tried to uncover the combination of predictor variables that would likely explain the failure of students across different math courses.

Method – The methodology follows a survey of 1200 DLSU-D students across the seven colleges of the university. Personality types were evaluated using the Myers-Briggs Type Indicator (MBTI), and mathematics anxiety was measured using the Abbreviated Math Anxiety Scale (AMAS). The machine learning implemented the framework of Probabilistic Graphical Models in Python (PGMPy) for data visualization and analyses. Predictions of possible grades were summarized, and the full Bayesian Network was established.

Results – Bayesian analyses have shown that the chances of failing a math subject are generally low for each year level. Personality variables conclude that college students with analyst roles have a higher probability of having a perfect 4.00 grade in a math subject than in an explorer role. Predicting the chances of failure between having or not having math anxiety seems almost no significant difference.

Conclusion – The subject, anxiety, consistency, and enjoyment variables are minimally enough to infer the probability of mathematics grades; hence all other variables can be ignored. There was a total of 8 math subjects with predicted probabilities of failing students over a total of 13.

Recommendations – Researchers recommend the use of other probabilistic graphical models aside from Bayesian Networks to verify and compare the joint probabilities between the variables of the study.

Research Implications – Providing comprehensive insights to properly accommodate at-risk students in each math subject will greatly help mathematics professors recalibrate their attention and teaching strategies.

Keywords – multivariate, Bayesian Networks, machine learning, graphical models, PGMPy

INTRODUCTION

Mathematics education has primarily been one of the most critical research concerns for educators. Predominant research in this field encompasses high school mathematics placement and students' attitudes toward the subject. While very few are seen to have been related to the college circumstances, mathematics semestral grades in college –

pertaining to different areas of mathematics subjects offered at the undergraduate level, could arguably impose varying levels of importance on the success of a college student. Depending on their respective course alignment, some might find mathematics as a very important factor in college success, and some might not. In many cases, the importance of predicting mathematics semestral grades has come into a great aid to assess college students who might probably need much more attention and guidance as the semester come to an end. With a number of factors, we can take into consideration, the use of machine learning algorithms – like the Bayesian Network, is seen to be useful and efficient in multivariate inferencing of the joint probabilities and independence between the mathematics grade predictors.

One particular variable predictor is the way a student's personality type or specifically behavioral temperament, based on David Keirsey's typology was able to determine their mathematical connections ability (Nurdiyah et al., 2018). Different students who have different personality types will hence manifest different intellectual abilities. It was found that students with rational temperaments tend to amplify all indicators of mathematics connection abilities. Another study was able to connect these personality factors to a measurable influence in predicting math anxiety and completion rates of mathematics students (Fuller et al., 2016). Students who have significant math anxiety tend to be incapable of completing the course successfully. Provision of interventions like mentoring or peer coaching the students are being concluded to overcome such anxiety barriers. In the same contextual passion, the relationship between math anxiety and mathematics performance has also been investigated through a meta-analysis (Zhang et al., 2019). The result confirms a robust negative correlation between math anxiety and mathematics performance. This means that the higher the math anxiety level of a student, the less their mathematics performance and achievement be. Subsequently, mathematics study habits and preferences also undermine a revealing impact on the mathematics performance and achievement of students (Verma et al., 2022). This recent study additionally implies that students with good mathematics study habits and preference exhibits higher mathematics achievement in their academic career.

Given these predominant studies in mathematics education, so far, we lack the specific measurement of mathematics performance concerning college students and their semestral mathematics grades in different college mathematics subjects. In connection to this, modeling the simultaneous interplay between multiple factors like personality type, math anxiety, and mathematics study preferences is much a great utility and advantage in predicting mathematics semestral grade. Bayesian Network machine learning algorithms are seen to be useful in explaining the causal relationship between multiple variables while simultaneously predicting the joint probability of multivariate outputs (Smail, 2016). Better recognition of at-risk students via sophisticated algorithms was recommended as it also affects the completion and graduation rate in college. Proper predictive assessments of mathematics semestral grades might as well help the faculties and professors to determine which students are at-risk and need immediate guidance in their early years of taking the course.

As a general overview, this study primarily wants to examine the model of relationships among the personality types and roles (using the Myers-Briggs Type Indicator), mathematics anxiety (using the Abbreviated Mathematics Anxiety Scale), and mathematics study habit preferences as the imperative variables that determine the mathematics semestral grade of a college student from De La Salle University – Dasmariñas (DLSU-D). Other specific demographic information is also being measured to yield important factors that might be relevant to the case study of DLSU-D students.

Research Questions

The study primarily wants to identify the answer to the following questions:

- What predictive probabilities of semestral grades might the students conclude given different factors such as their demographics, personality type or roles, mathematics anxiety, and mathematics study preferences?

- Which among these variables, within the overall interplay of all the factors, would be useful and minimal enough to predict the math semestral grades of a student?

- What variables or a combination of predictor variables would normally lead to a high probability of failure in different mathematics courses for college students?

LITERATURE REVIEW

Math Anxiety and the Abbreviated Mathematics Anxiety Scale (AMAS)

Psychologically, mathematics anxiety is expounded as a feeling of apprehension and doubts that interfere with the math cognitive ability and performance of a certain individual (Khasawneh et al., 2021). College students who agonize over mathematics anxiety tend to be less confident in their ability to do mathematical calculations, often leading them to hate the subject courses and tremendously limit their logical skills (Eispino et al., 2017). Whereas the Abbreviated Math Anxiety Scale is a predominant psychometric scale used in broad ranges of math anxiety research. It is a nine-item scale derived via dimensional reduction or factor analysis of the revised Math Anxiety Rating Scale (MARS) (Hopko et al., 2003). The AMAS uses a 5-point Likert Scale which is generally a valid math anxiety scale with very good psychometric properties, high reliability, and test-retest stability.

Personality Types, Temperaments, and Myers-Briggs Type Indicator (MBTI)

Personality types, according to the Swiss psychiatrist and well-known psychoanalyst Carl Jung (1971), are characterized by an individual's preference for a general attitude. The areas of preferences introduced by Jung are what he termed as "dichotomies" of the human cognitive functions: Extraversion (E) – Introversion (I), Sensing (S) – Intuition (N), and Thinking (T) – Feeling (F). Later research by Isabel Briggs, a practitioner of Jung's theory proposed to realize the Judging (J) – Perceiving (P) as the fourth dichotomy influencing personality types. Briggs, together with her daughter Katharine Myers have co-created the well-known personality type inventory Myers–Briggs Type Indicator (MBTI).

Personality temperaments on the other hand, also known as the personality roles, are the associated roles depending on the two shared traits dichotomy of the personality types: the Analysts share the Intuitive and Thinking traits; Diplomats share the Intuitive and Feeling traits; Sentinels share the Sensing and Judging traits; and the Explorers share the Sensing and Prospecting traits. These unique matching roles are primarily being used to pertain to less complicated psychoanalytical characteristics rather than involving all the possible combinations of personality types.

Bayesian Networks in Machine Learning

Bayesian Networks are graphical models from the idea of digraphs that represent the probabilistic dependencies of its variables via the notion of Directed Acyclic Graphs (DAG) or commonly termed acyclic digraphs. The vertices of these graphs serve as the variables while the arcs pertain to the links of the dependencies (Niedermayer, 2008). The information that is contained inside each arc link of a Bayesian Network comprises their conditional probability distributions (CPD). In this passion, each node shall have its own conditional probability table (CPT) that will show all the probabilities a vertex will have for the values of its dependencies. The major advantage of using the Bayesian Network lies in computing the probabilities, where one can easily figure out which among the vertices are relevant and which are not.

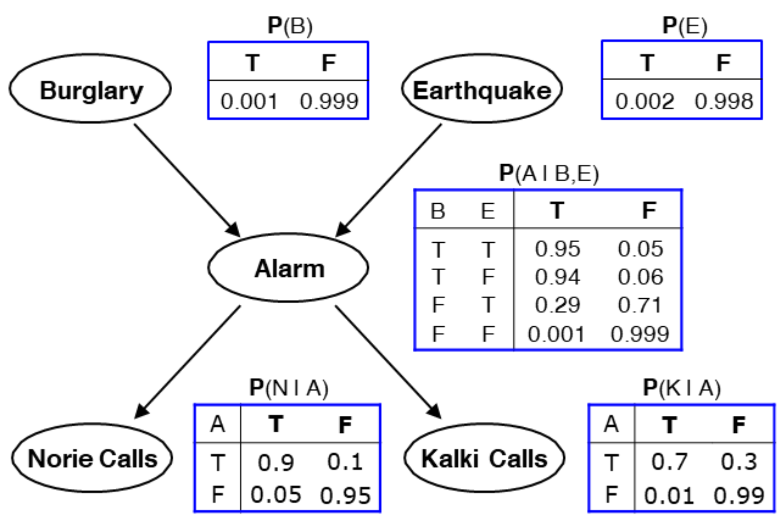

Figure 1. Bayesian Network Modelling of the Alarm System Example

As an example, suppose a certain company has an alarm system in anticipation of burglary incidences. Furthermore, the alarm system can occasionally get set off by sudden earthquakes and tremors. The company location has two adjoining neighbors, Norie and Kalki, who basically do not know each other. If they hear the alarm, they will call security, yet is not guaranteed. We want to represent the probability distribution of the events composed of Burglary (B), Earthquake (E), Alarm (A), Norie calls (N), and Kalki calls (K) as shown by the Bayesian Network model above.

In Figure 1, suppose we want to determine what is the probability that there is an alarm but neither a burglary nor an earthquake has occurred, and both Norie and Kalki called. The joint and probability is therefore given by the following solution:

$$\begin{aligned} \mathbf{P}(A \land \neg B \land \neg E \land N \land K) &= \mathbf{P}(A \mid \neg B, \neg E) \times \mathbf{P}(\neg B) \times \mathbf{P}(\neg E) \times \mathbf{P}(N \mid A) \times \mathbf{P}(K \mid A) \\ &= 0.001 \times 0.999 \times 0.998 \times 0.9 \times 0.7 \\ &= \mathbf{0.000628} \end{aligned}$$

Therefore, there is only a little chance of almost 0.06% for such an event to happen. With the causal relationships modeled out, the number of joint probabilities in the Bayesian Network can now be traced – making computations and memory use much more efficient. In addition, the information for the conditional probability distribution of each vertex is properly accounted for.

Bayesian Network modeling in machine learning has been the main focus of this paper since the fact that counterfactual analyses of the Bayesian predictions can well be formulated and understood. The network and associated predictions can also adapt quickly to changes in circumstances or changes in the underlying distributions. Moreover, Bayesian Network modeling via machine learning provides the rare opportunity to study the casualties of these events, allowing us to further study the educational dataset and uncover every probability information possible.

Related Studies

In the study from a university in Jordan, done by Smail (2016), the Bayesian Networks machine learning algorithm has been employed to analyze educational data and model relationships among the predictor variables such as personality types, math anxiety, study habits, and other demographics. In relation to the study being conducted, the study of Smail became the main literature that provided the idea of associating the psychology variables and demographics in modeling and predicting mathematics grades. The evident difference lies in the choice of the outcome or the dependent variable, as the main literature provides the multivariate associations of the factors in predicting mathematics anxiety, while this research study provides the multivariate associations of the factors in predicting the mathematics semestral grade of a different case study of participants – which are the college students from De La Salle University – Dasmariñas. On the other hand, Bayesian Networks are also useful as a technique for predicting whether academic advice should be given to a student or not. Using only academic factors – like semestral grades as the predictor of student performance, the likelihood of academic counseling can be lighted (Itoh et al., 2012).

In the same context of the study, Bayesian Networks can also be utilized in predicting exam scores of university students in a predefined course (Torabi et al., 2012). Like in the case of Azad University of Technology in India, Bayesian Networks modeling has been employed, but the machine learning implemented a different algorithm – the maximum likelihood algorithm. The study aimed with the same research objective as the proposed multivariate predictive modeling via Bayesian Networks – assisting the university officials and professors to predict student success and strategy management before the beginning of another semester. While in the case of another study in the Tamilnadu district of India (Ramaswami & Rathinasabapathy, 2012), the socioeconomic variables also played a great role in the prediction of a student's academic performance. The aim was to help their admission office to identify excellent students for allocating relevant scholarships based on the prior socio-economic predictor variables.

Other unique and novel applications of Bayesian Networks for educational purposes would include a study about developing and implementing an Intelligent Tutoring System (ITS) that is used to predict learning areas with which students had difficulties (Menor, 2018). In this study, two groups of students were tested – a control group and an experimental group. Bayesian Networks were employed to look for the probability of the learning difficulties from the experimental group. The results of the study were in turn used to present tutoring suggestions for the students with found difficulties.

Looking at the other facet of the context, predictive modeling via Bayesian Networks across a multitude of courses was also of pre-eminent response to the recommendations by an earlier study in Croatia (Oreski et al., 2019) where a significant limitation is the limited number of respondents – pertaining only to the case of BS Computer Science students. Since the proposed study of multivariate predictive modeling via the Bayesian Networks algorithm encompasses the multitude of DLSU-D students across different college courses, the validity of mathematics semestral grade predictions is seen to be justified.

These pieces of literature encompass a multitude of related studies dealing with the same aspect of predictive modeling via the Bayesian Networks algorithm in educational settings. Predominantly, combinations of different predictor variables being studied shall determine the robustness of the model.

METHODOLOGY

Area of Study

Generally regarded as the Premier University of Cavite, De La Salle University – Dasmariñas offers students an unconventional take on education, allowing them to learn in an environment surrounded by nature. Since DLSU-D has always been in the utmost passion of inculcating the Lasallian values of service, faith, and communion in every aspect of the student's academic life, this becomes the ground of study by this paper – allowing the machine learning applications on the case study of mathematics semestral grades of the college students to provide a comprehensive insight in properly accommodating students who are at risk in a certain math subject. This would in turn help the mathematics professors and faculties to recalibrate and refocus their attention on the most academically vulnerable students. The following are the mathematics college subjects offered by the Mathematics and Statistics Department (MSD) of the university: Advanced Calculus, Plane and Solid Mensuration, Number Theory, Biostatistics, Discrete Mathematics, Quantitative Methods, Calculus 1 for Engineers, Calculus 2 for Engineers, Introduction to Descriptive Analytics, Introduction to Predictive Analytics, Management Science with Computer Application, Statistical Analysis with Software Application, and Mathematics in the Modern World.

Data Gathering Procedures

A unique online questionnaire created through Microsoft Forms became the basis of the research for data gathering. The research questionnaire is furcated into 4 parts: the demographics part, the MBTI assessment, the AMAS assessment, and the study preferences part. Since the primary objective of the study is to characterize the combination of predictor variables that would normally lead to a high failure semestral grade for a certain mathematics subject, some other study preferences and demographic information were also measured to see the interplay between the predictor variables and the mathematics grade as the response variable. In addition, the consent form agreement and Lasallian student ID number were properly stipulated and requested. This is to ensure and verify that only valid Lasallian students were surveyed across the timeframe of online data gathering. A total of 1200 college students for the first semester of the school year 2022-2023, throughout the seven colleges of DLSU-D and across different courses and year levels were surveyed by the research.

Research Variables

The data on college students' identities were coded through their assigned student ID numbers to maintain confidentiality. The names and their corresponding Lasallian e-mails were also deleted as soon as the data were managed already. The research variables are grouped into four factors: demographic factor, personality factor, anxiety factor, and study habit factor. Variables considered under the demographic factor include proximity, year level, university experience, study advice, the respective college of the student, and the mathematics course subject they have taken. The personality factor includes personality role, personality type, sex, feeling, mathematics enjoyment, and challenging course enjoyment. On the other hand, variables under the anxiety factor comprise math anxiety, difficulty, difficult career, difficult task, and success field for college students. Lastly, the study habit factors encompass: study habit, consistency, impression, first step, forget, trouble, and stress variables. Students' math grades shall be the dependent or outcome variable of the analysis.

RESULTS AND DISCUSSIONS

Preliminary Descriptive Visualizations

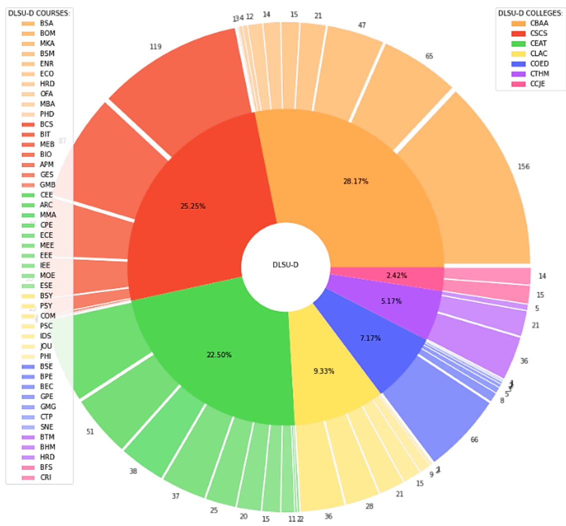

Figure 2. Pie Chart Visualization of the Respondents for each College and Course

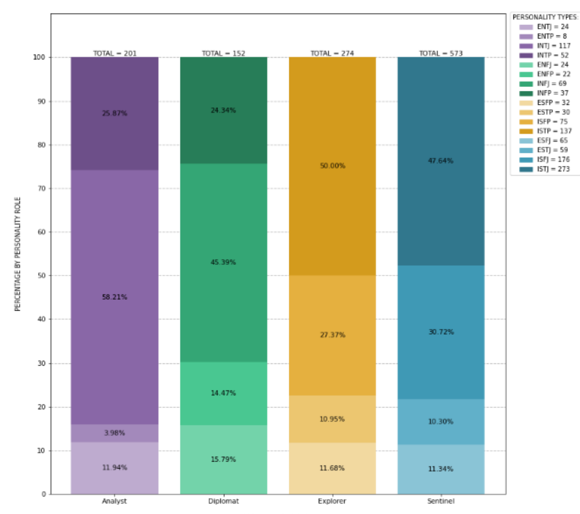

Figure 3. Stacked Bar Plot Visualization of Personality Roles and Personality Types

Figure 2 shows the percentage division of the respondents when aggregated via the college and course variables. The majority of the college students are from the College of Business Administration and Accountancy (CBAA), which makes up 28.17% of the total responses. While the college with the least number of respondents came from the College of Criminal Justice Education (CCJE), making up only 2.42% of the total surveyed students. Whereas if aggregated by their corresponding courses, it is seen that 156 of the CBAA students came from the BS Accountancy (BSA) program, with which the study had the greatest number of respondents by course. On the other hand, the programs with the least number of responses came from some unique individual representatives from the graduate studies of Doctor in Business Administration (Ph.D.), MS in Biology (GMB), and Certificate in Teaching Program (CTP).

In Figure 3, there were 573 sentinel students who constitute the largest group of personality roles of the surveyed students. The smallest group o,n the other hand, having a count of 152 students were of diplomat personality roles. Their respective personality types, as evaluated by the Myers-Briggs Type Indicator were also plotted via stacks. Of the sentinel college students, 43.64% or 273 were ISTJs, making them the most common type of personality among the DLSU-D students, followed by the ISFJs (with 176 students), and then the explorer ISTPs (with 137 students). The least the sentinels were ESFJs, which corresponds to roughly 11.34% of their count. While on the other hand, 45.39% or 69 of the diplomats were INFJs, which constitute the most common personality type of the role. The least common, with only 14.47% of the total counts of diplomats or equivalently were the ENFPs. For the analyst personality roles, the most common personality type is found to be the INTJs, which is 58.21% of the personality role or equivalently 117 students. The last of them, composed of 3.98% was the ENTPs. This makes them the least common personality type among the DLSU-D students with only 8 students out of the total surveyed college students being seen to have this particular personality type. Lastly, the 10.95% of the explorer personality role constitutes the ESTPs as their least fraction, which is equivalently 30 students in the count.

Bayesian Networks Approach

Since we want to model the Bayesian Network that describes the interplay between the predictor variables and the outcome variable, the methodology approach of this study divides the Bayesian predictive modeling into 4 parts: via the interplay of the demographic variables, personality variables, anxiety variables, and the study habit variables. The study aims to predict the possible semestral grades the Lasallians students might conclude given such factors. The machine learning processes implemented the framework of Probabilistic Graphical Models in Python (PGMPy). Other specific results were not explicitly included. Only the most important and selected inferences were extracted and hereby generalized by the discussion. The succeeding figures (Figures 4 to 7) of Bayesian Network models are learned by the machine via the following lines of code:

```
# IMPORTING PYTHON LIBRARY
from pgmpy.models import BayesianModel

# MULTIVARIATE PREDICTIVE MODELLING DATA FRAME
MultivariatePredictiveModelling =
pd.read_csv("MultivariatePredictiveModelling.csv")

# DEMOGRAPHIC VARIABLES BAYESIAN NETWORK MODELLING
DemographicNetwork = nx.DiGraph()
DemographicNetwork.add_edges_from([("COLLEGE", "COURSE"), ("COLLEGE",
"YEAR\nLEVEL"), ("YEAR\nLEVEL", "ADVICE"), ("COURSE", "MATH\nSUBJECT"),
("YEAR\nLEVEL", "EXPERIENCE"), ("PROXIMITY", "EXPERIENCE"), ("MATH\nSUBJECT",
"MATH\nGRADE")])
plt.show()

# DEMOGRAPHIC VARIABLES PROBABILISTIC PREDICTION
DemographicModel = BayesianNetwork([("COLLEGE", "COURSE"), ("COLLEGE", "YEAR
LEVEL"), ("YEAR LEVEL", "ADVICE"), ("COURSE", "MATH SUBJECT"), ("YEAR LEVEL",
"EXPERIENCE"), ("PROXIMITY", "EXPERIENCE"), ("MATH SUBJECT", "MATH GRADE")])
DemographicModel.fit(MultivariatePredictiveModelling)
DemographicModel.get_cpds()
```

```
# DEMOGRAPHIC MODEL CHECK
DemographicModel.check_model()

# DEMOGRAPHIC CONDITIONAL PROBABILITY ON YEAR LEVEL
DemographicGradeInference = VariableElimination(DemographicModel)
PROBABILITYYearLevel = DemographicGradeInference.query(variables = ["MATH GRADE"], evidence = {"YEAR LEVEL": "Freshman"})
print(PROBABILITYYearLevel)
```

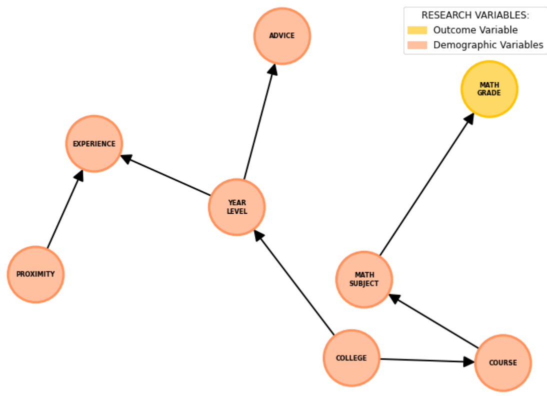

Figure 4. Bayesian Network Model to Predict Mathematics Grade via the Demographic Variables

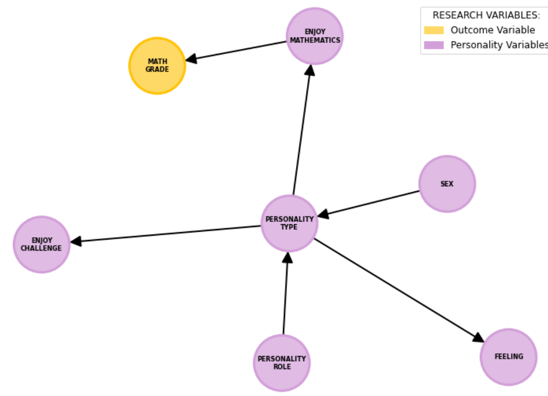

Figure 5. Bayesian Network Model to Predict Mathematics Grade via the Personality Variables

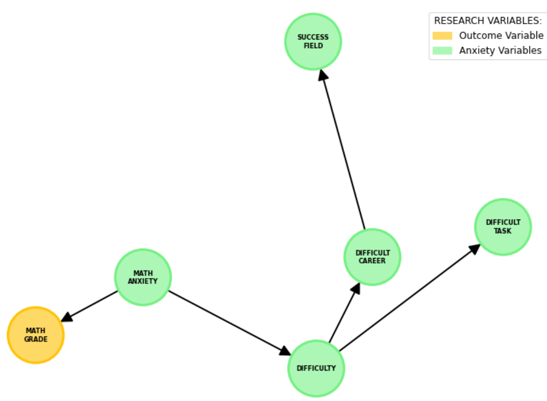

Figure 6. Bayesian Network Model to Predict Mathematics Grade via the Anxiety Variables

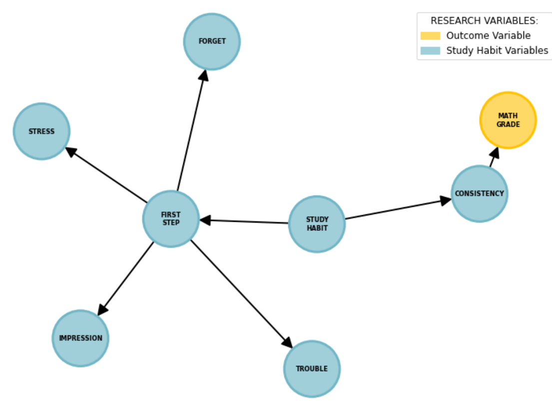

Figure 7. Bayesian Network Model to Predict Mathematics Grade via the Study Habit Variables

We can play along with the code by changing the desired commands, depending on the variable and joint probabilities being calculated. The following table summarizes some selected inferences from the study to answer the research questions:

Table 1. Probabilistic Prediction of Math Grade given Different Prior Information

Math Grade	P(Math Grade | Year Level = Freshman)	P(Math Grade | Year Level = Senior)	P(Math Grade | Personality Type = ENTP)	P(Math Grade | Personality Type = ISTP)	P(Math Grade | Math Anxiety = No)	P(Math Grade | Math Anxiety = Yes)	P(Math Grade | Study Habit = Regularly)	P(Math Grade | Study Habit = Before)
0.00 (Failed)	1.86%	1.18%	1.95%	1.58%	1.57%	1.58%	1.36%	1.64%
1.00 (60% - 63%)	2.02%	1.32%	0.98%	1.75%	0.90%	1.82%	1.71%	1.76%
1.25 (64% - 67%)	1.01%	0.74%	0.50%	0.92%	0.45%	0.95%	0.69%	0.98%
1.50 (68% - 70%)	2.16%	1.56%	2.07%	1.92%	2.02%	1.91%	1.88%	1.93%
1.75 (71% - 73%)	1.88%	1.26%	1.63%	1.67%	1.57%	1.67%	1.40%	1.74%
2.00 (74% - 76%)	2.32%	2.04%	2.50%	2.16%	1.57%	2.21%	1.90%	2.24%
2.25 (77% - 79%)	3.32%	2.71%	3.17%	3.08%	3.15%	3.08%	3.14%	3.07%
2.50 (80% - 82%)	5.86%	5.23%	5.65%	5.67%	5.17%	5.70%	5.32%	5.76%
2.75 (83% - 85%)	7.78%	7.81%	5.55%	7.76%	6.52%	7.84%	7.00%	7.96%
3.00 (86% - 88%)	11.21%	11.54%	8.85%	11.18%	8.31%	11.39%	11.38%	11.11%
3.25 (89% - 91%)	10.57%	10.75%	10.15%	10.50%	8.99%	10.62%	10.07%	10.62%
3.50 (92% - 94%)	17.14%	17.79%	16.02%	17.26%	15.73%	17.37%	19.08%	16.75%
3.75 (95% - 97%)	15.80%	16.58%	17.31%	16.08%	20.22%	15.77%	16.36%	16.01%
4.00 (98% - 100%)	17.05%	19.48%	23.67%	18.47%	23.82%	18.09%	18.73%	18.44%

And as provided by Table 1, given that we know a certain student is a freshman, the predicted probability of failing his or her math subject course is just 1.86%. This hence would conclude that he or she will pass with a probability of 98.14%. On the other hand, when compared to the predicted probability when a DLSU-D student is a senior, the probability of failing is reduced to 1.18%, and hence passing is increased to 98.82%. Additionally, accounting for the probability of getting the highest grade point of 4.00 in DLSU-D increases by 2.43% between being a freshman to a senior.

Given that we know a certain student is an ENTP, the predicted probability of failing his or her math subject course is just 1.95%. It follows that he or she will pass with a probability of 98.05%. On the other hand, when compared to the predicted probability when a DLSU-D student is an ISTP, the probability of failing is reduced to 1.58%, and hence passing is increased a little bit to 98.42%. Additionally, an ENTP is 5.20% greater in the chance of having a 4.00 grade in comparison to an ISTP.

Knowing that a college student does not have evident math anxiety, we can predict with a 23.82% probability of having a perfect 4.00-grade point in a math subject course. This is 5.73% higher than having math anxiety. However, when comparing the chances of failing a math subject, the state of having or not having math anxiety appears to be almost identical.

The predicted probability of failure for those DLSU-D students who study regularly in a given semester is lower than for those who study just before the summative exams. Similarly, if a student studies regularly, he or she has a 98.64% chance of surviving the semester, compared to a 98.36% chance if he or she does not. With these, the study wants to determine which among the variables in the interplay of all 4 factors would be useful enough to predict the mathematics semestral grade of a student – hence we want to identify the Markov Blanket of the collective Bayesian Network.

By running the following lines of code where *"ArcList"* is the list of all the directed connections of the Bayesian Network, we have determined that consistency, enjoying mathematics, math anxiety, and math subject variables are the blankets.

```
MarkovBlanket = BayesianNetwork(ArcList)
MarkovBlanket.get_markov_blanket("MATH GRADE")
```

Figure 8. Markov Blankets for the Math Grade Variable

The colored nodes in Figure 8, depict the subgraph of the full Bayesian Network with which knowing only about the consistency of a student in studying at least two hours, if he or she enjoys math, has math anxiety, and the math subject he or she is taking, would be enough to predict the math grade. This shall be the minimal number of information we need to infer something about the mathematics grades. By these results, the summary table below shows the probabilities of failing based on these blanket variables.

Table 2. Probabilities of Failing given the Math Subject, Math Anxiety, Consistency, and Enjoy Mathematics Prior Information

Math Subject	Math Anxiety	Consistency	Enjoy Mathematics	**P**(Fail)
G-MATH100	Yes	No	Yes	9.60%
	No	No	Yes	9.18%
	Yes	Yes	No	8.34%
	Yes	No	No	3.85%
S-EMAT001	Yes	No	No	18.35%
	No	Yes	No	13.33%
	Yes	Yes	No	6.05%
	Yes	No	Yes	5.86%
	No	Yes	Yes	4.02%
S-EMAT002	Yes	No	No	2.97%
S-ITPC223	Yes	No	No	0.39%
	No	Yes	No	0.22%
	Yes	No	Yes	0.17%
	No	Yes	Yes	0.10%
S-MATH112	Yes	No	No	3.17%
	Yes	No	Yes	1.62%
S-MATH200	Yes	Yes	Yes	1.98%
	No	No	Yes	1.28%
S-MATH201	Yes	No	No	8.34%
S-MATS314	Yes	Yes	No	0.53%
	No	No	No	0.32%
	No	Yes	No	0.31%

As depicted in Table 2, integrating the prior information from these four variables into the Bayesian Network model, we can conclude that the 47.61% chance of failing is attributed to those students who take S-EMAT001 or Calculus 1 for Engineers. Simply put, there is an almost 50% odd of failing basically if a student is an Engineering student in DLSU-D. The highest among these combinations accounts for an 18.35% predicted chance of failing from the combination of having math anxiety, being not consistent in studying at least 2 hours a day, and not having to enjoy the subject. On the other hand, taking S-EMAT001 without math anxiety and knowing that a student is consistent in studying still accounts for a 13.33% chance of failing the subject. Additionally, failing G-MATH100 or Mathematics in the Modern World accounts for an odd of 30.97%. Predicted chances of failing mathematics subjects Quantitative Methods (S-ITPC223) and Advanced Calculus (S-MATS314) are seen to be minimal. Overall conclusion, there are predicted chances of failing 8 out of the 13 math subjects in DLSU-D. The rest which was not explicitly shown in Table 2, hence have no predicted probabilities of failure.

CONCLUSIONS

The multivariate prediction of grades via Bayesian Network models is seen to be helpful in dealing with large educational data sets. Collectively understanding the interplay of different demographic, personality, anxiety, and study habit factors in a causal-relational passion, predictions, and inferencing via the framework of Bayesian Network machine learning algorithm is seen to be efficient to a college setting with a multifaceted plethora of student background and attitudes. As elucidated by the study, results were able to predict a different range of probabilities in getting a certain grade point in mathematics across different math subject courses offered by De La Salle University – Dasmariñas.

Analyses have shown that in comparing the mathematics grades between freshman (1st year) and senior (4th year) DLSU-D college students, students who have taken their math courses as seniors are more likely to pass than those who were freshmen when they have taken the subject. Meanwhile, when comparing the ENTP and ISTP personality types, an ENTP has a greater chance of having a 4.00 grade in comparison to an ISTP person. To add, it can also be validated through the machine learning process that students who do not enjoy mathematics but feel very interested in talking to their math professors are predicted to have a higher chance of having a perfect 4.00-grade point in their math subject. In predicting the probable state of a college student given that he or she failed in math, it is found that the student is more likely to say that it is much more difficult to solve mathematical problems than to write an essay, and both being a mathematician and a philosopher is way much more difficult rather than being a mathematician or a philosopher alone. Results also show that students who have 4.00 math grades are ironically predicted to be both forgetful and stressed by their math subjects. Probabilistic inferences state that a college student is more likely to fail if he or she takes or is currently taking Calculus 1 for Engineers, has math anxiety, is consequently not consistent in studying, and does not enjoy the subject course more than the other mathematics subject courses in DLSU-D. Predicted probabilities of failing a general math subject Mathematics in the Modern World are also evident. There are a total of 8 mathematics subjects with predicted probabilities of a failing student over a total of 13.

RECOMMENDATIONS

Since the research was not able to derive the possible connection between each of the demographics, personality, anxiety, and study habits, the researchers propose to study the relationships of the variables across the factors themselves. Additionally, the study can further be refined by adding more factors or variables to the model – like the physiological attributes of the college students, their socio-economic status, relational dispositions, and maybe some preferential attitude towards things that might likely be relative in explaining their academic deportment towards college Mathematics.

Lasallian researchers may also continue updating the dataset learned by the machine, hence adding more information gathered from DLSU-D students to further have more accurate probabilistic inferences. Given that the resulting Bayesian probabilities are not enough to justify the significance of the prediction, the researchers also recommend the use of other probabilistic graphical models, aside from the Bayesian Networks, to verify and compare the joint probabilities between the variables of the study. Or if possible to be continued, justification can be done by conducting verification procedures after the current semester of the school year 2022-2023 in DLSU-D. Any other hypotheses concluded by the research can further be studied through some correlational or structural validation.

IMPLICATIONS

DLSU-D, as a university that caters to a diversity of undergraduate students, has always been in support of regular professor-student interaction. Providing comprehensive insight to properly accommodate students who are at risk in a certain math subject would greatly help the math professors recalibrate their attention. Furthermore, applications of the Bayesian Networks algorithm in educational data provide flexibility and efficiency rather than manually computing the probability of the multivariate factors. As studies in mathematics education emerge, more and more data needed to be analyzed and considered. Novel applications of different predictive modeling and machine learning algorithms are becoming requisite to study large data sets.

ACKNOWLEDGEMENT

The researchers would like to extend their deepest gratitude to the Computer Studies Department (CSD), and Mathematics and Statistics Department (MSD) of De La Salle University – Dasmariñas, especially to the respective department chairs Ms. Josephine Eduardo and Ms. Sharon Lubag for the utmost support in making this research possible. To the unconditional efforts of Dr. Ma. Theresa Christine Valdez, Dr. Lea Maderal, and Mr. Edwin Bunag for endorsing the online research survey to their students, all the best to them!

Sincere thanks are expressed to the University Research Office, especially to Dr. Jemerson Dominguez, chair of the DLSU-D Ethics Review Committee for approving the conduct of the study. Thank you for the guidance and support.

To the college student governments who willingly assisted in disseminating the survey link – Mr. Mark Joshua Emmanuel Visleño of COEDSG, Ms. Hannah Kathrina Morallo of the CTHMSG, Ms. Kristine Joyce Nava of CLACSG, and Ms. Lizette Lei Bleza of CCJESG, we thank you very much.

Ultimately, the warmest appreciation is being extended to all Lasallian students who patiently allotted time to answer the research survey. May your support of this endeavor be a reflection of the true Lasallian identity in aiding for the betterment of the least, the lost, and the last people of the community. May this scientific venture be ample evidence of how we teach the mind, touch the heart, and eventually try to transform the lives of our fellow Lasallians. Sincerely, in Saint La Salle!

DECLARATIONS

Conflict of Interest

The authors conjointly declare that there is no conflict of interest.

Informed Consent

All college participants of the study were informed of the purpose and data to be collected upon answering the online survey questionnaire. A consent form agreement and proper ethics approval were requested for the duly conduct of the study. Lasallian student ID numbers were also requested to ensure and verify that only valid Lasallian students were surveyed across the timeframe of data gathering. Some identities of the respondents like their names and Lasallian e-mails were deleted as soon as the targeted number of participants approached, hence before the data analyses commenced.

Ethics Approval

The DLSU-D University Research Office (URO) and the Institutional Ethics Review Committee (IERC) accepted and approved the conduct of the study.

REFERENCES

Eispino, M., Pereda, J., Recon, J., Perculeza, E., & Umali, C. (2017). Mathematics Anxiety and Its Impact on the Course And Career Choice of Grade 11 Students. *International Journal of Education, Psychology and Counselling*, 2(5), 99-119

Fuller, E., Deshler, J., Darrah, M., Trujillo, M., & Wu, X. (2016). Anxiety and Personality Factors Influencing the Completion Rates of Developmental Mathematics Students. *HAL Archives Ouvertes*. https://hal.archives-ouvertes.fr/hal-01337892

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS). *Assessment,* 10(2), 178–182. https://doi.org/10.1177/1073191103010002008

Itoh, K., Itoh, H., & Funahashi, K. (2012). Forecasting Students' Grades Using a Bayesian Network Model and an Evaluation of Its Usefulness. *IEEE Xplore*. https://doi.org/10.1109/SNPD.2012.84

Jung, C. G. (1971). Psychological types: The collected works, volume 6. London, UK: Routledge and Kegan Paul.

Khasawneh, E., Gosling, C., & Williams, B. (2021). What impact does maths anxiety have on university students? *BMC Psychology*, 9(1). https://doi.org/10.1186/s40359-021-00537-2

Menor, J. (2018). Development and Implementation of an Intelligent Tutoring System Using the Bayesian Network in Teaching C# Programming Language. JETRB4 Journal of Engineering and Technology Research, 6(2), 1–12.

Niedermayer, D. (2008). An Introduction to Bayesian Networks and Their Contemporary Applications. In: Holmes, D.E., Jain, L.C. (eds) Innovations in Bayesian Networks. *Studies in Computational Intelligence*, vol 156. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-85066-3_5

Nurdiyah, N., Suyitno, H., & Junaedi, I. (2018). Mathematical Connections Ability Based on Personality Types in Conceptual Understanding Procedures Model. *Unnes Journal of Mathematics Education Research*, 7(1), 9–17. https://journal.unnes.ac.id/sju/index.php/ujmer/article/view/21282

Oreski, D., Konecki, M., Pihir, I. (2019). Predictive Modelling of Academic Performance by Means of Bayesian Networks. *Economic and Social Development: Book of Proceedings*:435-41.

Ramaswami, M. & Rathinasabapathy, R. (2012). Student Performance Prediction Modeling: A Bayesian Network Approach*. International Journal of Computational Intelligence and Informatics*. Vol. 1, No.4

Smail, L. (2016). Using Bayesian networks to understand relationships among Math anxiety, genders, personality types, and study habits at a university in Jordan. *Journal on Mathematics Education*, 8(1), 17-34.

Torabi, R., Moradi, P., & Khantaimoori, A. R. (2012). Predict Student Scores Using Bayesian Networks. *Procedia - Social and Behavioral Sciences*, 46, 4476–4480. https://doi.org/10.1016/j.sbspro.2012.06.280

Verma, N., Chitkara, M., & Malhotra, R. (2022). Statistical analysis of study habits in mathematics achievement. *AIP Conference Proceedings*. https://doi.org/10.1063/5.0080558

Zhang, J., Zhao, N., & Kong, Q. P. (2019). The Relationship Between Math Anxiety and Math Performance: A Meta-Analytic Investigation. *Frontiers in Psychology*, 10. https://doi.org/10.3389/fpsyg.2019.01613

Authors' Biography

Norie Neil C. Acosta is currently a 4th year BS Applied Mathematics student at the College of Science and Computer Studies of De La Salle University – Dasmariñas, Cavite, Philippines. He was a recipient of the Gawad Talino 2019 Excellence in Mathematics Award by the Mercury Drugs Foundation. He is a consistent academic scholar, financial assistance scholar, city scholar, and provincial scholar – currently maintaining a Summa Cum Laude Latin honor. As an Applied Mathematics student, he has a prospective inclination to truth and certainty of pure mathematics, statistics, data science, machine learning, and artificial intelligence. Norie Neil is currently under his thesis writing in spatiotemporal data mining of dengue epidemiology.

Christian Kalki M. Lamadrid is currently a 3rd year BS Computer Science student under the College of Science and Computer Studies of De La Salle University – Dasmariñas, Cavite, Philippines. He is taking the Intelligent Systems track and focusing on developing practical AI applications in software and robotics. He is proficient in software engineering and web development and has a background in computer engineering. Christian Kalki is adept in different programming languages and has an eye for their applications in different scientific research and studies.

Mr. Rolando Barrameda is an Asst. Professor at the Computer Studies Department of De La Salle University - Dasmarinas, Cavite, Philippines, and has been teaching for nearly 16 years. He earned his Master's in Information Technology and continues to pursue his Ph.D. in Computer Science, currently working on his Dissertation with a specialization in Deep Learning, Machine Learning, and Artificial Intelligence. He authored and co-authored research papers that were published in different Scopus Indexed International Journals. He was a trainer and a speaker who delivered several local and international paper presentations. A certified Microsoft Office Specialist and Certified Microsoft Technology Associate.