

Bayesian Network report

CSC4025Z : Artificial Intelligence

Hamza Amir
amrham001@myuct.ac.za
University Of Cape Town
Cape Town, South Africa

Adeeb Gaibie
gbxade002@myuct.ac.za
University Of Cape Town
Cape Town, South Africa

Dennis Hammerschlag
hmmnden001@myuct.ac.za
University Of Cape Town
Cape Town, South Africa

1 INTRODUCTION

South Africa has a high number of secondary school students who fail to complete their education, particularly those who are vulnerable due to factors such as poverty or lack of family support. To better understand the dropout rate, we consulted educational professionals from our former high schools. Predicting a student's likelihood of dropping out could be valuable in providing support to vulnerable students before it is too late.

In this project, we aim to develop a Bayesian network to predict the likelihood of a student dropping out of school. Bayesian networks are a type of probabilistic graphical model that can represent uncertain relationships between nodes. By using a Bayesian network, we can model the dependencies between the various factors that contribute to a student dropping out of school.

There are a number of extra lesson initiatives around South Africa aimed to help vulnerable students. Our second part of the project involves building a tool/decision network that will be used by school professionals and policy makers to determine:

- whether the student needs extra school workshops

that subsequently brings about [more students graduating] with the [reduction in the high school dropout rate for students].

This project can have significant implications for education policymakers and researchers interested in understanding the factors that contribute to students dropping out of school.

2 PROBLEM ANALYSIS

The problem of children dropping out of school is a significant issue that has far-reaching consequences for both individuals and society and it has been aggravated since the Covid-19 pandemic. According to News24, the pandemic may have caused three times as many students to drop out of school, with around 750,000 learners having left school in May 2021, compared to an expected 230,000 pre-pandemic. [11]

We consulted an educational professional who has been actively involved in non-profit educational workshops over the past few decades. Various high school teachers were also contacted in terms of obtaining more knowledge. In addition to utilizing expert knowledge; multiple studies, including ones specific to South Africa, were used to build the Bayesian network.

We identified main factors that contribute to dropout rates including (but not limited to) school motivation [4], family support [15], financial support [13]. The identification of students who are at risk of dropping out is crucial for schools to intervene and prevent this outcome. Therefore, the development of effective tools, such as Bayesian networks, to predict the likelihood of a student dropping out of school is of great importance.

2.1 School Motivation Factors

Attendance, academic performance and home internet access are significant predictors of student dropout rates. Whilst this is just one aspect of a student's livelihood, it plays a major role in the likelihood of students dropping out. Some data from [2] provides insights on the relation between attendance and academic performance on school motivation as a whole. Stats SA also evidently has good data in this area.

2.1.1 Attendance. Attendance is a critical factor that affects the likelihood of a student dropping out as it reflects their engagement and commitment to their education. Regular attendance indicates active participation in learning and better understanding of the course material, while also providing opportunities for peer interaction and support from instructors. Learning losses due to reduced attendance of students is currently a major challenge for the schooling system, which will require long-term, concerted efforts to rectify. [7] Statistics on children in South Africa has relevant statistics on attendance that are utilised in this model. [10]

2.1.2 Academic Performance. Academic performance has evidently been seen to be negatively linked to the likelihood of a student dropping out. [5] The better a student performs in their curriculum system, the less pressure they feel academically which in turn doesn't make them feel pressurized negatively enough to want to drop out. Academic performance amongst South African high-schools are generally not the best [14]. But the pass rate for 2022 was celebrated to be 80% by the Department of Education [3]. Whilst no exact data for academic performance in South Africa is available, from the aforementioned data-sets, we can get a brief understanding of the overall distribution of it. Further, academic performance can be improved with the help of extra classes. A study into the effectiveness of extra classes provides statistics about their efficacy.[9]

2.1.3 Internet. Internet access is crucial for a student's motivation in school. Without it, they may lack access to valuable educational resources, including online textbooks, research articles, and interactive learning platforms. Furthermore, limited communication with teachers and peers due to a lack of internet access can result in disengagement from the learning process. Overall, not having internet access can be a significant obstacle to academic success. According to the General Household Survey conducted in 2016 ([17], it has been evident that around 60% of South African households had at least one member of the family who had access to or used the Internet either at home, work, place of study or at cafes. But only roughly 10% had access to the internet at home.

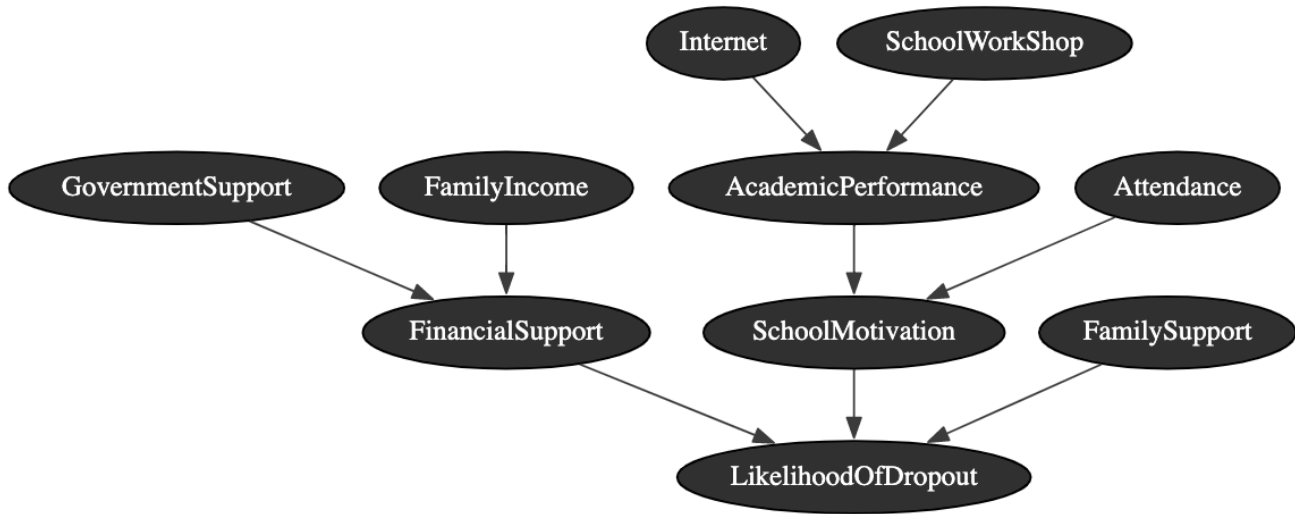


Figure 1: Bayesian Network Model - "SchoolWorkShop" is a decision node in the Decision Network but modelled as a uniform random variable in the Bayesian Network as it affects Academic Performance.

2.2 Family Support Factors

Family support factors are instrumental in shaping a student's academic trajectory [1], serving as a vital pillar of support during times of academic hardship. These include parental support and the state of the student's home environment. When families support you, you can help students develop their motivation and self-confidence in the classroom [20].

Parents/guardians tend to be students' first point of contact when they need assistance diagnosing and troubleshooting the vast majority of their errors in their subjects. Parental involvement is crucial in mitigating the risk of dropping out. [12] Active participation in school activities, such as parent-teacher conferences and volunteering, conveys the significance and value of education to the child. [6]

Challenging home environments, including poverty, unstable housing, domestic violence, and substance abuse, can create stress and chaos for students, hindering their ability to focus on academics, leading to frequent absences, falling behind, and ultimately, dropping out. A student's home environment has to be conducive for learning in order to aid them in the learning process and reduce the likelihood of them dropping out. [8] The General Household survey conducted in 2019 by Stats SA has good data on the nature of households that students find themselves in. [18]

2.3 Financial Support Factors

Financial support plays a critical role in reducing the probability of a student dropping out. Insufficient financial resources can pose a myriad of challenges, including difficulty in affording basic school supplies, transportation costs, and living expenses. Additionally, financial strain may force students to take up part-time jobs or other responsibilities, thereby impeding their focus on academic pursuits. Conversely, financial aid in the form of scholarships, grants, and

other forms of assistance can ease the burden of these challenges and enable students to fully immerse themselves in their education, thereby mitigating the risk of dropping out.

2.3.1 Government Support. The extent to which a student can receive financially government support depends on the government's policy regarding student funding and the requirements therein. Usually government funding for secondary students is less than what it would be for tertiary education. Think Impact has good statistics about funding for students. [19]

2.3.2 Family Incomes. Family incomes can have a significant impact on a student's likelihood of dropping out of school. Students from low-income families are at a higher risk of dropping out due to financial difficulties, including the need to work or help support their families. These students may also lack access to resources that could help them succeed academically, such as tutoring or technology. Additionally, students from low-income families may be more likely to attend underfunded schools with fewer resources, which can further hinder their academic success and motivation to stay in school. The wealth disparity seen in South Africa is one of the largest in the world, and data of that can be seen here [16] or at the World Bank. [21]

3 DECISION NETWORK MODEL

Drawing upon insights from our educational professional and our extensive research into the key factors outlined in the preceding section, we have formulated a Bayesian structure, as depicted in Figure 1 before. Subsequently the decision network can be seen in Figure 2 above. As mentioned previously from various data sources in alignment with the advice received from educational professionals, academic performance of students is pivotal to their likelihood of dropping out. Educational professionals usually take the stance

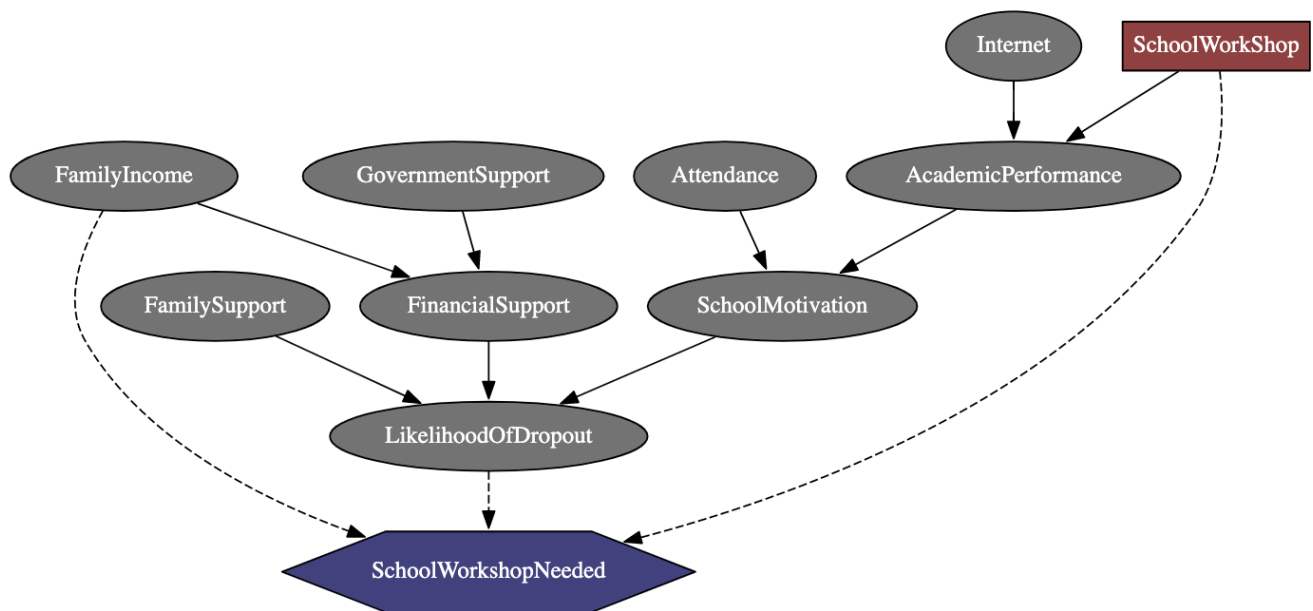


Figure 2: This shows the Decision Network Model, where the grey, red, and blue nodes represent chance, decision, and utility nodes, respectively.

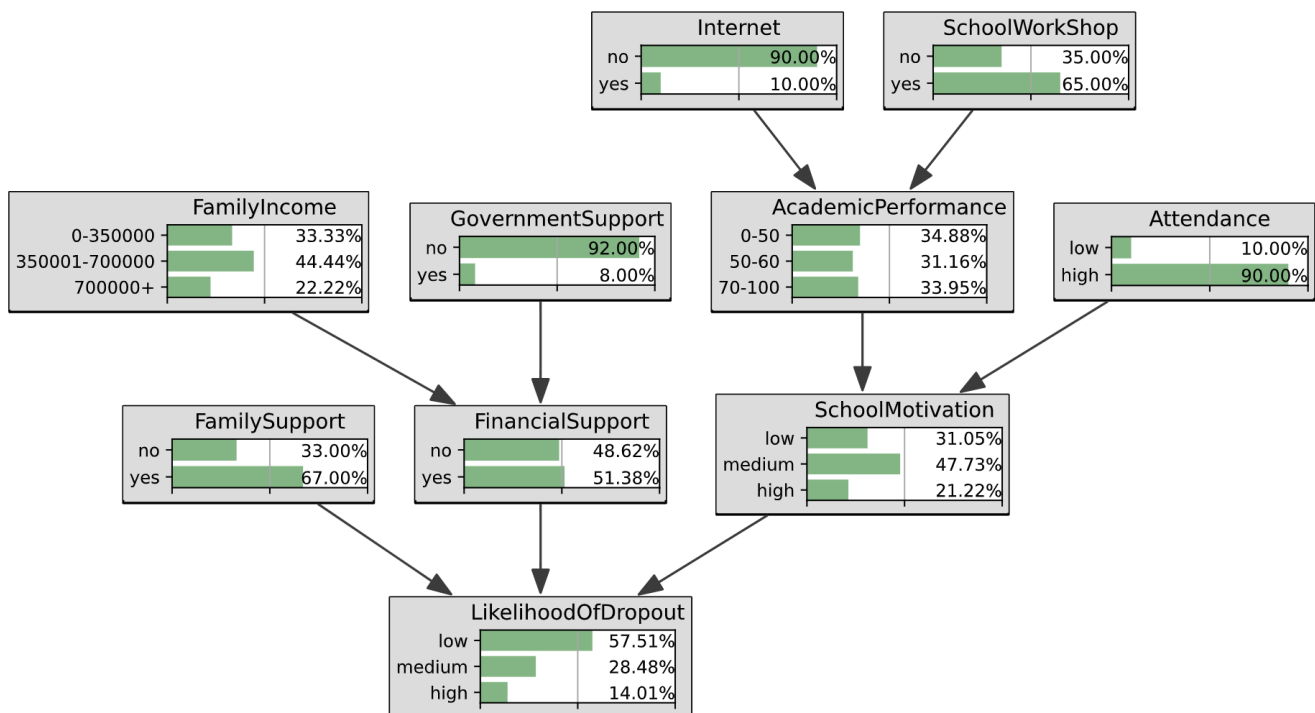


Figure 3: Posterior Probabilities for our Bayesian Network

of allowing access to school workshops for all, but what's needed

is to be able to target specific academically-deficient performing student groups that need it the most.

Thus, we have decided to include the utility node denoted as "SchoolWorkShopNeeded" which measures a student's attending these non-profit workshops. Factors influencing a student's applicability to attend these workshop are as follows:

3.1 Family income

The non-profit workshops are primarily aimed in helping the underprivileged. Therefore, the student's family income needs to be considered as ideally students from lower income households should be have a higher preference of attending these workshops. Therefore the utility on lower income households is weighted higher.

3.2 Dropout likelihood

These workshops have the goal of lowering dropout rates. Therefore, students with that are more vulnerable to dropping out are more applicable for attending the workshops. Thus, the utility is weighted higher for students more vulnerable to dropping out.

3.3 Workshop

As there is a limited amount of resources, such as voluntary teachers available, not everyone can attend these workshops. Therefore to account for this, attending the workshop itself is negatively weighted against the utility. This stops the network from recommending the workshop to everyone.

The factors mentioned above all influence the utility, however weightings of each factor need to be considered. There is no concrete data values on how each factor should be weighed. With consultation with our educational professional, we were able to gauge which factors are most important. It was deduced that the workshops would favour towards having any underprivileged student with higher grades over a privileged student with lower grades. Therefore, the weighting of family income is higher than dropout likelihood.

4 MODEL TESTING AND EVALUATION

Utilizing the Bayesian Network depicted in Figure 1, it is possible to perform inference in the absence of evidence by leveraging the posterior probabilities that are illustrated in Figure 3. Additionally, the decision networks can be employed in a similar fashion, as demonstrated in the following use cases:

Case 1 - Scenario 0

As seen on the following page, by not providing the school workshop to the student results in the highest utility. The optimal decision therefore would to recommend the student to not attend the school workshop. This particular use case serves to illustrate the exceptional ability of our model to arrive at optimal decisions by relying solely on probabilistic information, even in the face of incomplete knowledge about present circumstances.

Case 2 - Scenario 1

In this case, the evidence points us in a clear direction towards a decision. The lack of home internet access has a negative impact on academic performance, while low attendance levels have a similar

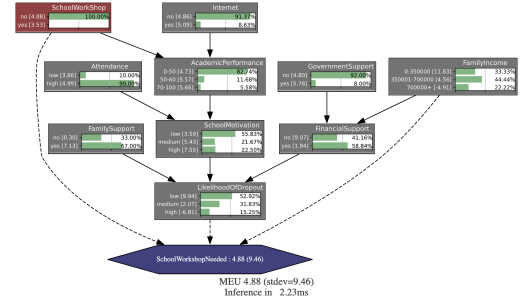


Figure 4: Scenario 0 with no evidence applied.

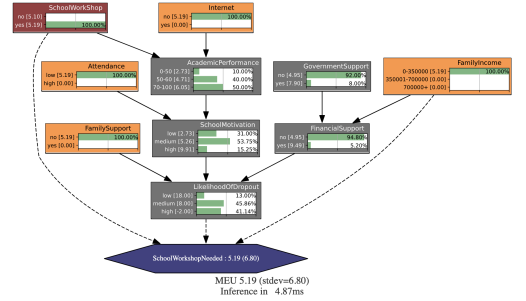


Figure 5: Scenario 1 with some evidence applied.

effect on school motivation. Furthermore, family income in the range of 0-350,000 and lack of family support have a negative influence on financial aid and the likelihood of dropout, respectively. Considering these factors, the utility node of a student in need of a school workshop is influenced by family income, likelihood of dropout, and the decision to attend the workshop itself. Thus, our model recommends that the student attend the workshop, which aligns with the intuitive evidence. This use case demonstrates the effectiveness of our model in making decisions that are consistent with our expectations based on the available evidence.

Case 3 - Scenario 2

Within this use case, we invert the evidence set to test the effectiveness of our Decision Network model in recommending whether or not a student should attend school workshops. The nodes of home internet access and family support are both set to "yes," along with high attendance and a family income of 700,000+. Our model correctly recommends that the student does not need to attend the workshop, which aligns with the intuitive evidence. This demonstrates the remarkable ability of our model to make decisions that are in line with our expectations based on the available evidence.

Case 4 - Scenario 3

In the final use case, we deliberately set ambiguous evidence to test the Decision Network model's effectiveness in recommending whether a student should attend school workshops. The nodes of home internet access is set to "yes" alongside with attendance as "high" and home support set to "no". The family income is set to the middle value of "350 001-700 000". These factors can be seen

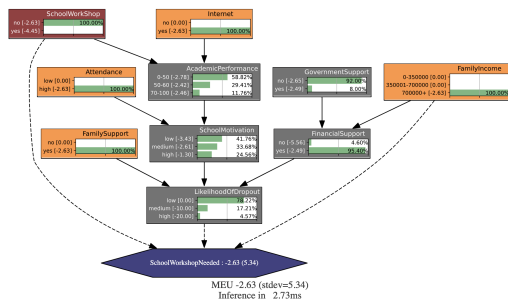


Figure 6: Scenario 2 with some evidence applied.

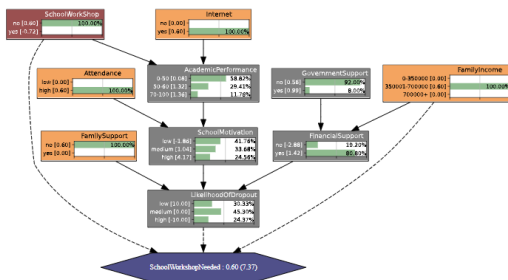


Figure 7: Scenario 3 with some ambiguous evidence applied.

to positively and negatively affect the likelihood of a student to drop out, but it depends on the implications of the interactions of these nodes. The derived optimal decision would thus be to recommend that the student does not attend the school workshop. Our model demonstrates the unique ability to make decisions based on probabilities, even when humans may not fully understand the implications of the evidence provided.

5 CONCLUSION

The Bayesian network model that was produced had the aim to provide an accurate estimation of the uncertainties involved in a high school student dropping out and not completing their education. Although it is not entirely comprehensive, there are many other factors that influence the likelihood of a student dropping out, we have evidently captured the main factors identified in the literature. Further, we have extended the model to a decision network in order to identify the most suitable candidates for non-profit workshops and have shown the model to perform good decision approximations. Future research could model the network on other locales, with relevant influential factors, as the decision network has been designed to be extensible. For these reasons, we consider the BDN to be successful and useful to the intended user group.

REFERENCES

- [1] ASSOCIATION, A. P. Parent engagement in schools. <https://www.apa.org/pi/lgbt/programs/safe-supportive/parental-engagement/default.aspx>, n.d.
- [2] BALFANZ, R., . B. V. The importance of being in school: A report on absenteeism in the nation's public schools.
- [3] BUSINESS TECH. 2022 matric pass rate climbs to 80.1% in south africa, 2019. Accessed : 2023-06-04.
- [4] COVINGTON, M. V. Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology* 51, 1 (2000), 171–200. PMID:

- 10751969.
- [5] DEPARTMENT OF BASIC EDUCATION. Learner progress and achievement study, 1998. Accessed : 2023-05-30.
- [6] DEPARTMENT OF BASIC EDUCATION. Family influences on early grade reading, 2017. Accessed : 2023-06-02.
- [7] DEPARTMENT OF BASIC EDUCATION. The covid-19 pandemic, enrolments, dropping out and attendance explained, 2022. Accessed : 2023-05-28.
- [8] DEPARTMENT OF BASIC EDUCATION. Update on the review of the policy on home education, 2022. Accessed : 2023-06-02.
- [9] DUC, L. T. Do extra classes improve cognitive test scores, Dec. 2012.
- [10] HALL, K. Statistics on Children in South Africa : School attendance, 2022. Accessed : 2023-06-03.
- [11] MARECIA, D. As many as 750,000 children have dropped out of school during the pandemic, 2021. Accessed : 2023-05-27.
- [12] NATIONAL EDUCATION EVALUATION DEVELOPMENT UNIT. Building a bridge between school and home: a sample of best practices how to improve parental involvement in schools, 2018. Accessed : 2023-06-01.
- [13] ORION, H. C., FOROSUELO, E. J. D., AND CAVALIDA, J. M. Factors affecting students' decision to drop out of school. *SLONGAN* 2, 1 (Aug. 2014), 16.
- [14] REDDY, V., BERG, S., JANSE, D., JANSE VAN RENSBURG, D., AND TAYLOR, S. Educational outcomes: Pathways and performance in south african high schools. *South African Journal of Science* 108 (12 2011).
- [15] RUMBERGER, R. W., AND ROTERMUND, S. *The Relationship Between Engagement and High School Dropout*. Springer US, Boston, MA, 2012, pp. 491–513.
- [16] SA-TIED. Extreme inequalities: The distribution of household wealth in south africa, 2020. Accessed : 2023-06-07.
- [17] STATS SA. General household survey, 2016. Accessed : 2023-06-06.
- [18] STATS SA. Families and parents are the key to well-being of children, 2019. Accessed : 2023-06-04.
- [19] THINK IMPACT. Scholarship statistics, 2021. Accessed : 2023-06-03.
- [20] WAIRIMU, M. J., MACHARIA, S. M., AND MUIRU, A. Analysis of parental involvement and self-esteem on secondary school students in kieni west sub-county, nyeri county, kenya. *Journal of Education and Practice* 7 (November 2016), 82–98.
- [21] WORLD BANK. South africa : Income, 2023. Accessed : 2023-06-08.