University of Cape Town

CSC4025Z

Bayesian Networks Assignment

AI Assignment 1

Author:
Stephan Maree
Benjamin Harvey
Racquel Dennison

Student Number:
MRXSTE013
HRVBEN001
DNNRAC003

October 2, 2024

Stephan Maree
Benjamin Harvey
Racquel Dennison

Contents

1 Background

Higher tertiary education encompasses all formal post-secondary education, including both public and private universities. For years, it has been argued that tertiary education plays a critical role in fostering economic growth, reducing poverty, and promoting shared prosperity. A higher education qualification not only benefits the individual but also significantly contributes to societal well-being. University graduates tend to be more environmentally conscious and more engaged in civic activities. Additionally, their higher earnings increase tax revenues, which in turn stimulates economic growth. [1]

Despite the clear benefits of higher education, South African education statistics reveal that only 22% matriculants attend university.[2] Various factors influence a student's likelihood of attending university. The objective of this report is to explore these factors and evaluate whether a student is likely to apply to university, using a Bayesian network.

The data used to compute probabilities in the Bayesian network is synthesized from an Indonesian context which considers social factors that influence students' interest and the ultimate outcome of applying to university. This network can be extended to a South African context.

By utilising our Bayesian network, educators can identify the key factors affecting a student's decision to apply to university and implement strategies to address these challenges. For example, if they see that students' grades are a major contributor to applying to university, educators can intervene by offering ways of improving students' grades. This tool aims to maximize university applications by identifying and addressing the root causes which affect students proceeding to university. Understanding these factors will allow targeted interventions and raise awareness, ultimately helping more students pursue higher education.

2 Problem analysis

The potential factors that were considered in our Bayesian network were as follows :

1. Students parents went to university
2. Students parents age
3. Average salary of parents

[1] https://www.worldbank.org/en/topic/tertiaryeducation
[2] https://www.southafricanmi.com/education-statistics.html

4. School accreditation (public or private school)
5. School type (Academic or vocational)
6. Interest in applying to university
7. Average grades

The data set can be found https://www.kaggle.com/datasets/saddamazyazy/go-to-college-datasethere. Below is an explanation of each of the points focused in the data set:

- School Type: There are two main school types focused on in the data. The one is academic which is orientated on academic studies and research skills, while the other is vocational schools. Vocational schools focus on applied and practical skills such as mechanics and hairdressing. Research shows that students who attend vocational skills are less likely to succeed in university so would not even apply [5].
- School Accreditation: A is the best, followed by B and C. While school accreditation doesn't measure student continuity to college, college admissions view school accreditation as one of the student's qualifications [3].
- Residence: Whether the student grew up in an Urban area or a Rural area. Residents of urban areas are found to be more likely to obtain a degree from a university than rural residents [4].
- Parents Age: Parents age, we would assume that more senior parents have a higher salary to assist their children through university [7].
- Parents Salary: Parents' salary would impact students going to university as funding is a contributing factor to which schools they go to and what their average grades are like.
- Average Grades: Averaged high school grades. This is scored out of 100.
- Parents in university: If parents went to university, this can impact a student's interest and motivation to go to university [2].

3 Decision Network model

The final Decision Network model aimed to capture the key variables and relationships involved within students' environments when applying for university. The structure was carefully designed to reflect dependencies between these variables with

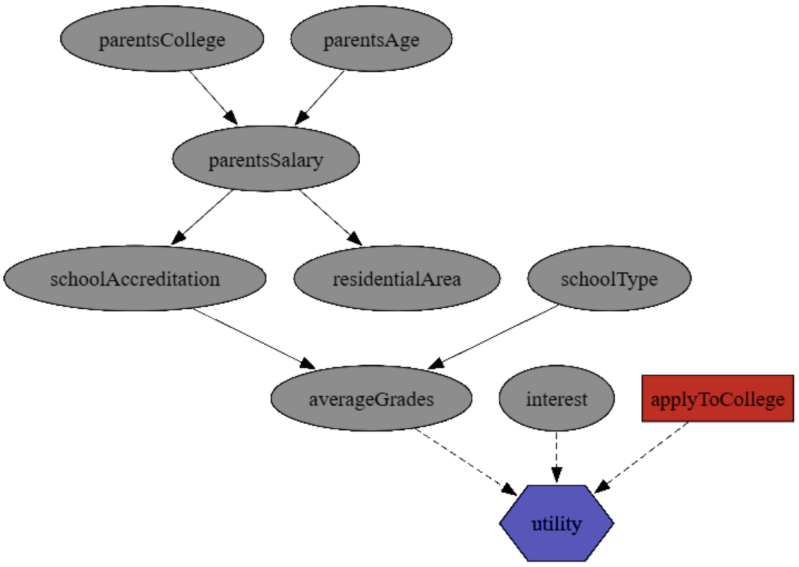

Figure 1: Decision network

backed-up references. Figure 1 shows a graphical representation of our decision network.

3.1 Relationship between nodes

3.1.1 Parents' education and age affect their salary

Research shows that parents who attended university and obtained a higher degree tend to have higher-paying jobs, confirming a correlation between education and salary. This relationship is crucial because it highlights the long-term economic impact of higher education on individuals' earning potential [7]. Furthermore, age is also an important factor that affects salary, as older individuals often accumulate more industry experience, which leads to higher earnings [7]. In our model, the parents' education and age nodes are treated as parent nodes to the salary node, meaning that both education and age directly influence salary. This structure reflects the logical flow that education increases skill levels, and age reflects experience, both of which contribute to higher earnings.

3.1.2 Parents' salary affects the school accreditation

The parents' salary node is also linked to the school accreditation node. Higher parental income often enables families to afford better educational institutions for their children, which may be accredited and offer a higher standard of education.

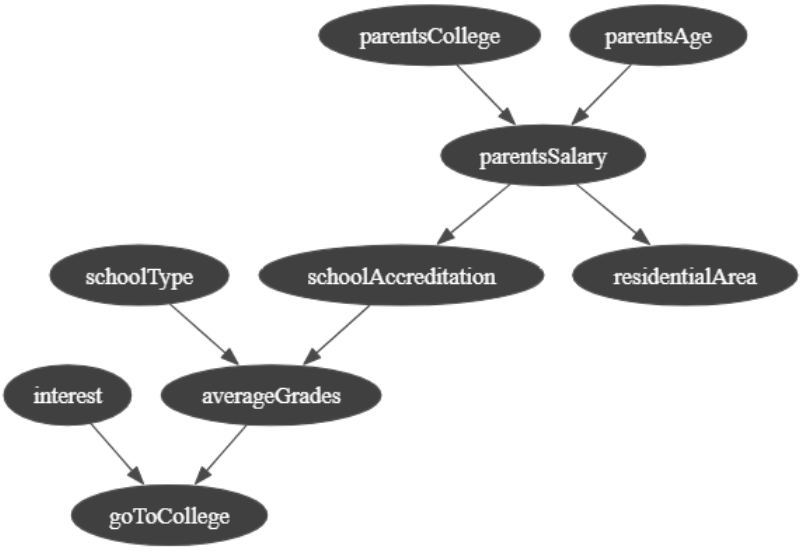

Figure 2: Bayesian Network

Studies indicate that financial resources give families access to schools with better reputations and accreditation standards [1]. Put simply, more educated and richer parents can provide a "better" environment for their children.

3.1.3 School accreditation and type affects grades

In our model, the average grade is affected by both the school accreditation and the type of school. The of school has a major impact on the student's grades as higher accredited schools often provide better teaching and attention to student's needs [6]. Studies have shown that vocational schools are often chosen by students who struggle or fail in academic schools [5], so it is reasonable to expect students who attend to have lower grades on average.

3.1.4 Student's Interest Affects their chance of going to college

If a student has a strong desire to continue their studies they will have a higher probability of taking the necessary steps—such as applying to universities and preparing for entrance exams— to increase their chances of enrollment. These students are more likely to try and overcome financial, academic, and logistical barriers in order to achieve their goal of going to college.

4 Model testing and evaluation

Using the Bayesian network, we can make various inferences based on which variables we observe. For example, we can marginalise out the unknown variables and obtain the tables for P(goToCollege|Interest):

P(College|Interest=Yes)

College	Probability
Yes	0.4992
No	0.5008

P(College|Interest=Maybe)

College	Probability
Yes	0.5551
No	0.4449

P(College|Interest=No)

College	Probability
Yes	0.4660
No	0.5340

We can use the Bayesian network to determine if the school students attend has an influence on applying to college. This is shown in Table 1.

Table 1: P(GoToCollege|SchoolAccreditation)

P(College|SchoolAccreditation=B)

College	Probability
Yes	0.4833
No	0.5167

P(College|SchoolAccreditation=A)

College	Probability
Yes	0.5186
No	0.4814

From this, we see that students who attended a school of type A would more likely apply to university as opposed to those attending a school of type B.

Another inference that can be made is investigating whether or not the parents going to college has an effect on the student's odds of going to college. We can compute the following conditional probability tables:

P(College|ParentsCollege=Yes)

College	Probability
Yes	0.5003
No	0.4997

P(College|ParentsCollege=No)

College	Probability
Yes	0.5003
No	0.4997

The fact that the probabilities are the same suggests that the goToCollege node is independent of the parentsCollege node. Knowing whether or not your parents went to college seems to not provide any additional predictive power as to whether or not the student will go to college.

Table 2: Utility function for the decision network

Apply	Interested	Grades	Utility
Yes	Yes	Above Average	100
Yes	Yes	Below Average	40
Yes	No	Above Average	70
Yes	No	Below Average	-100
Yes	Maybe	Above Average	80
Yes	Maybe	Below Average	0
No	Yes	Above Average	-100
No	Yes	Below Average	0
No	No	Above Average	-50
No	No	Below Average	100
No	Maybe	Above Average	20
No	Maybe	Below Average	0

Our model was extended to a decision network that allows the user to determine whether or not they should apply to college. The utility function can be found in Table 2.

This function can be used to compute recommendations for students based on which of the variables they have available.

For example, if a student has no interest in college, the expected utility is -23.62 for applying, and 32.61 for not, meaning that the model recommends that they don't apply.

Should no variables be observed at all, the expected utility for applying is 20.46 and for not applying is 2.02. This means that the model generally recommends applying to college, unless given evidence that contradicts this. The inferences used in this are shown in Figure 3.

5 Conclusion

Using this Bayesian network model we were able to illustrate the relationships between key factors involved in college enrollment. This analysis highlights that working on key variables such as improving school accreditation, reaching higher grades and fostering a greater interest in higher education will lead to higher enrollment rates.

Ultimately, understanding and building towards a society in which students are encouraged and able to pursue higher education is critical not only for individual

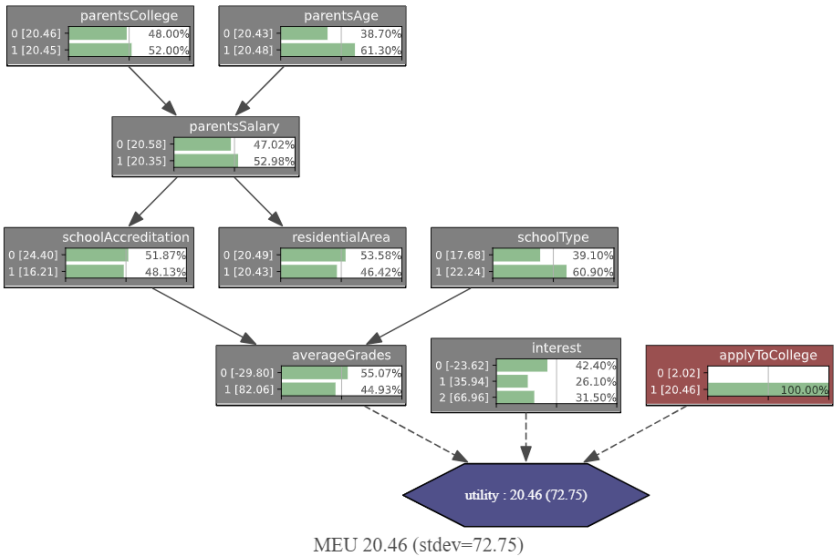

Figure 3: Inferences used to calculate the Maximum Expected Utility given no evidence variables.

success but for a broader nationwide advancement. By addressing and gaining a better understanding of the root causes behind low enrollment, we can move towards the goal of higher education for students.

References

[1] Harmon C. O' Sullivan V. et al. Chevalier, A. The impact of parental income and education on the schooling of their children. *IZA J Labor Econ 2, 8 (2013).*

[2] Huesmann LR. Dubow EF, Boxer P. Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill Palmer Q (Wayne State Univ Press)*, 2009 Jul;55(3):224-249. doi: 10.1353/mpq.0.0030. PMID: 20390050; PMCID: PMC2853053.

[3] Robert Hill. Does ieb make the grade? alternative testing methods and educational outcomes: The case of the ieb in south africa. *DEVELOPMENT POLICY RESEARCH UNIT*, November 2019.

[4] K.B. Newbold and W.M. Brown. The urban-rural gap in university attendance: Determinants of university participation among canadian youth. *Journal of Regional Science, 55(4), pp.585–608. doi:https://doi.org/10.1111/jors.12197.*, 2015.

[5] Todd Pugatch. Safety valve or sinkhole? vocational schooling in south africa. *IZA Journal of Labor & Development*, 2014.

[6] Nick Taylor. What's wrong with south african schools. In *What's Working in School Development Conference, JET Education Services, Cape Town*, 2008.

[7] J.C. van Ours. Age, wage and productivity. *Social Science Research Network*, 2010.