Education Success Rates based on Various factors.

Naiema O. Elsaadi

Northwest Missouri State University, Maryville MO 64468, USA
s556619@nwmissouri.edu

Abstract. This project will explore the significance that specific demographic, geographic, psychographic, and financial factors have on a student's academic performance by looking at data sources that show the academic performance of several students in relation to the mentioned factors. This project will use data analysis, data cleaning, and predictive models to determine how a student will do based on these factors

Keywords: Data Cleaning · Student Success · Education · Data Analysis · Predictive Models

1 Introduction

This paper focuses on Education. Education is often the most important pillar of success in the world today. Success in the classroom can be a good indicator of future success in employment, as well as be the beginning of a life of financial fulfillment. The data problems that this project hopes to analyze, are the factors that lead to student success in schools vs factors that lead to failure in schools. This is significantly important in today's world, because education is proven to be very important in the future success of most people. This Project will attempt to highlight the main factors that lead to student success and student failures, so that people can be more aware of them.

1.1 Why This Domain was Chosen

This domain was chosen, because success rates in education is an important topic, and unraveling the factors that lead to higher success rates in education is incredibly important. Whether those factors are geographical, demographical, psychographical, or financial, it is important to see the affects that these factors have on how well a student does in school.

1.2 Data Sources

The dataset that will be used for this project will be found through Kaggle, a popular platform for datasets. The datasets used will be ones that provide information on Student Performance in schools, GPAs, ACTs, Drop out rates, as

well as information on financial situation of students, geographic, demographic, and psychographic information on students and their families. This project may also contain data from government education websites and research institutions for this data.

2 Project Implementation Phases

This project will have the following implementation phases:

- Phase 1: Acquire Data source
- Phase 2: Data Preprocessing and Cleaning
- Phase 3: Exploratory Data Analysis
- Phase 4: Designing Predictive Models
- Phase 5: Building Predictive Models
- Phase 6: Model Evaluation
- Phase 7: Final Model Selection
- Phase 8: Results and Discussion
- Phase 9: Conclusion
- Phase 10: Limitations and Future Work

3 Key Components and Limitations

Some of the key components of my approach are the Data source, and the preprocessing steps. The data source will be really important for this project, because it contains all of the information that we need to make our conclusion about the factors that leads to success in education. The preprocessing and Cleaning up of the data is also a really important component of this project. There are numerous exceptions and outliers all around the American Education system, that may skew our data or invalidate it. Because of this, cleaning up the data will be really important to make sure we get rid of these outliers.

Some of the limitations of this project are also related to the data sources. This project may face some limitations in relation with the quantity and quality of the sources acquired. This could be caused by the data sources being limited to specific areas, or specific types of schools. Another limitation that this project could be affected by is that the data can not account for the internal factors within a student, such as mental health issues, that could affect a student's performance in school.

4 Data Collection

This data source is a source that contains information about academic success and some variables that are connected to it. This sourceis found on Kaggle and it is called "Predict Student's Dropout, Academic Success". The format this dataset is found in is in the format of a CSV File. No data scraping techniques

were used for gathering this data apart from downloading in onto a computer. This dataset contains lots of very interesting and important information that relates to the goals of this project. The most important attributes from this dataset that will be used in this project will be Mothers/Fathers qualification, Mothers/Fathers Occupations, Courses. Education Special Needs, Scholarships, Displaced, and Age of Enrollment. There are not many other data extraction details related to this project apart from the general preprocessing and cleaning of the data.

5 Data Preprocessing and Cleaning

6 Exploratory Data Analysis

7 Designing Predictive Models

8 Building Predictive Models

9 Model Evaluation

10 Final Model Selection

11 Results and Discussion

12 Conclusion

13 Limitations and Future Work

14 References

□

References

1. https://www.nwmissouri.edu/login/
2. Christenson, S.L., Rounds, T., Gorney, D.: Family factors and student achievement: An avenue to increase students' success. School Psychology Quarterly **7**(3), 178 (1992)
3. Knuth, D.: Knuth: Computers and typesetting, http://www-cs-faculty.stanford.edu/ uno/abcde.html
4. Kumar, K.N.: Predict students' dropout, academic success. https://www.kaggle.com/datasets/naveenkumar20bps1137/predict-students-dropout-and-academic-success?select=dataset.csv (2021), (accessed: 2023-10-22)