Fairness in AI: A Numerical Analysis on Student Performance in Secondary School

Danielle Sim
Applied Data Science, MS
University of Southern California
simd@usc.edu

Saurabh Jain
Applied Data Science, MS
University of Southern California
sjain681@usc.edu

Abstract

The primary goal of the project is to investigate possible bias in a dataset and examine how such bias affects a prediction model. The dataset at hand comes from the University of California, Irvine's Machine Learning Repository, called the 'Student Performance Data Set' [2] The prediction model goal is to predict students' final math grades which is on a 0 to 20 scale. We aim to understand the bias in data by investigating selected protected features such as sex, and the bias in prediction based on training multiple linear regression models to predict final math grades. Finally, we will explore some ways to mitigate bias in our predictions and explore how different definitions and implementations of fairness affect the results.

1 Project Domain and Goals

As mentioned before, in addition to the primary goal of this project of investigating possible bias in this dataset and examine its effects on prediction modeling, other objectives include implementing methods that may mitigate bias and determine if a fairer model can be achieved. The 'Student Performance Data Set' comes from secondary education data of two Portuguese schools and contains 649 observations or students and 33 attributes including sex, age, family size, address (urban or rural), parent demographic information, and other information such as student health, absences, and extra curricular activities. The diverse set of features that include demographic, social, and school related information for each student will allow for an extensive analysis on potential confounding variables and correlations with the prediction outcomes on final math scores. Several ways to mitigate bias will be explored, such as data augmentation, fairness through unawareness, group fairness or statistical parity in which the distribution of good outcomes are the same between two groups of a protected feature, conditional statistical parity, equalized odds, and equality of opportunity. We will also explore methods to achieve fairness on the individual level such as implementing counterfactual fairness, and lastly fairness on the subgroup level. For each implementation of these definitions of fairness, prediction models to predict final math scores will be constructed and to examine and compare accuracy and fairness.

Other potential areas of this project include converting the outcome variable (a numerical grade) into a binary variable (High, Low) and running logistic regression models, repeating the analysis plan mentioned above and compare results. Secondly, this data set contains another table for the same group of students but for their grades in a Portuguese language class. We also are interested in repeating this analysis for these grades as well, as it could be interesting to compare results between a STEM class and a non-STEM class for the same group of students, and determine how biases may stay the same or differ across classes.

Preprint. Under review.

2 Related Work

Studies have been done to determine features that can predict student academic performance irrespective of their level of study. One such study is "Using Data Mining to predict Secondary School Performance" [2]. It confirms the conclusion found in Predicting Students' Performance in Distance Learning Using Machine Learning Techniques [5]: student achievement is highly affected by previous performances. Nevertheless, an analysis comparing the best predictive models has shown that there are other relevant features such as: school related (e.g. number of absences, reason to choose school, extra educational school support), demographic (e.g. student's age, parent's job and education) and social (e.g. going out with friends, alcohol consumption) that best predict academic performance.

Meanwhile, a case study of Machakos teachers college portrays that students admitted with high grades are expected to outperform those admitted with lower grades but this was not the result; groups of students who had received C+ in KCSE (Kenya Certificate of Secondary Education) performed approximately the same as those who had scored a C across subjects [1]. This indicates that the student admission grade into the college does not count in their performance in the Primary teacher education curriculum. Similarly the students' age and gender are not contributory factors in their academic achievement according to this study. Rather, focus and preparedness determine good performance regardless of these demographic attributes.

Various studies have found that students from the least affluent socio-economic groups tend to perform less well than their more affluent peers [9]. Finally, personal characteristics such as sex and ethnicity are also known to influence academic performance. Though the present study does not focus on ethnicity, significant differences in performance and participation have been documented between ethnic groups. Another study confirms that by treating student performance in the core knowledge courses as an early warning signal, faculty, administrators, and students might be able to identify students who could benefit from early intervention and help them increase their probabilities of academic success in business studies [4]. While these studies' primary focuses have been on predicting academic performance with respect to specific features such as demographic attributes or previous academic history, in this study we will be searching for any source of bias in predicting academic performance with respect to protected features such as student age.

References

[1] Mutuku Christopher and Kiilu Redempta. Influence of demographic factors on academic performance among primary teacher trainees - a case study of machakos teachers college. *International Journal of Educational Studies*, 3(1):07–11, 2016.

[2] Paulo Cortez and Alice Silva. Using data mining to predict secondary school student performance. *EUROSIS*, 01 2008.

[3] Linda Green and Gul Celkan. Student demographic characteristics and how they relate to student achievement. *Procedia - Social and Behavioral Sciences*, 15:341–345, 2011. 3rd World Conference on Educational Sciences - 2011.

[4] Mehdi Kaighobadi and Marcus Allen. Investigating academic success factors for undergraduate business students. *Decision Sciences Journal of Innovative Education*, 6:427 – 436, 07 2008.

[5] Sotiris Kotsiantis, Christos Pierrakeas, and P. Pintelas. Predicting students'performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18:411–426, 01 2004.

[6] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning, 2022.

[7] Maliha Nasir. Demographic characteristics as correlates of academic achievement of university students. *Academic Research International*, 2(2):400, 2012.

[8] Shubham Sharma, Yunfeng Zhang, Jesús M. Ríos Aliaga, Djallel Bouneffouf, Vinod Muthusamy, and Kush R. Varshney. Data augmentation for discrimination prevention and bias disambiguation. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020.

[9] Tamara Thiele, Alexander Singleton, Daniel Pope, and Debbi Stanistreet. Predicting students' academic performance based on school and socio-demographic characteristics. *Studies in Higher Education*, 41(8):1424–1446, 2016.