A Fairness-Centered Comparison of Machine Learning Models for Student Retention

Deborah D. Kanubala[*]   KANUBALA@CS.UNI-SAARLAND.DE
Saarland University, Department of Computer Science, Saarbrücken, Germany

Alidu Abubakari[*]   A.ALIDU@CAMPUS.UNIMIB.IT
University of Milano-Bicocca, Department of Computer Science, Milan, Italy

Abstract

Machine learning (ML) is increasingly used to support decision-making in high-stakes domains such as healthcare, finance, and criminal justice. Particularly, in education, ML applications have grown substantially, ranging from personalized learning systems to early warning systems for student success. These applications have demonstrated measurable improvements in intervention timing and resource allocation. While ML offers a wide range of benefits in education, it also risks reinforcing societal biases. In this study, we leverage different ML algorithms to predict students at risk of dropping out, hence enabling timely interventions. We further assess the fairness of these models using different group fairness notions such as demographic parity, equality of opportunity, false positive rate and accuracy parity. Our main contribution is to assess which notion of fairness is best suited for the education domain. This research aims to align the potential of ML with ethical considerations, contributing to fairer and more effective educational interventions.

Keywords: Machine Learning, Student Dropout Prediction, Fairness, Education

1. Introduction

Machine learning (ML) is increasingly used to support decision-making in high-stakes domains such as healthcare, finance, and criminal justice, and is now gaining traction in education (Nieto et al., 2019; Nieto Acevedo et al., 2019; Al-kmali et al., 2020). While ML systems offer significant benefits, including improved efficiency and early detection capabilities, numerous studies have exposed their potential to perpetuate or amplify social biases, particularly those related to race (Buolamwini and Gebru, 2018; Sweeney, 2013) and gender (Dastin, 2022). This raises pressing ethical concerns, especially in education, where ML-driven decisions can directly influence students' academic trajectories and access to support. A central challenge in education is identifying students at risk of dropping out. Timely detection is crucial for effective intervention and long-term student success (Aulck et al., 2016). ML presents a promising solution by enabling early warning systems that help institutions act before it is too late (Perdomo et al., 2023). However, the application of ML in this context is not without risk. Historical examples such as the UK's exam grading algorithm (Chen et al., 2023) and the COMPAS recidivism tool (Angwin et al., 2022) show how seemingly neutral algorithms can result in discriminatory outcomes when fairness is not rigorously assessed.

---

[*] Equal contribution

Despite increased attention to fairness in ML, defining and enforcing fairness remains challenging. Over 20 different fairness definitions exist, which generally fall into two categories: i) *individual fairness*, which involves treating similar individuals similarly (Dwork et al., 2012) and ii) *group fairness*, which is ensuring outcomes are independent of group membership. A common group fairness notion is *demographic parity*, which requires that the probability of a favourable outcome be the same across demographic groups. However, prior work has shown that not all fairness notions can be simultaneously satisfiable (Miconi, 2017).

In this study, we investigate the use of group fairness metrics to audit ML models used for predicting student dropout. Using the dataset provided by Realinho et al. (2022), we aim to assess both the predictive accuracy and fairness of several ML classifiers. Our main goal is to identify which fairness metrics are most appropriate for the educational context. Hence, the contributions of this work are twofold:

1. Evaluate the predictive performance of three popular ML models (Logistic Regression, Decision Trees, Random Forest) for student dropout prediction, providing empirical comparisons across multiple evaluation metrics.

2. Conduct a comprehensive fairness audit using four group fairness metrics to determine their relative importance and suitability for educational decision-making contexts.

This paper is structured as follows: *Section 2* reviews related work and outlines the research gap. *Section 3* details our methodology, including dataset preprocessing, model selection, and fairness metrics. *Section 4* presents empirical findings and we provide a discussion of our results and their implications in *Section 5*. *Section 6* concludes and outlines directions for future research.

2. Related Work

Definitions and Perspectives on Student Retention. The definition of student retention varies among institutions and contexts, leading to a lack of consensus in the literature (Heritage et al., 2022). However, according to Robbins et al. (2004), it is generally understood as the length of time a student remains enrolled in a particular program or institution. Thus, student retention, in the context of education, refers to an institution's ability to retain students and keep them actively engaged in their academic programs until they graduate. Student retention is closely linked to graduation outcomes and academic performance in higher education, making it a concern for educational institutions, educators, and students alike (Friedman and Mandel, 2011). Student retention is crucial for justifying the sustainability and growth of educational programs and institutions (Kampf and Teske, 2013). Retention challenges include social, emotional, and systemic factors, as well as specific issues like persistent behaviour problems and the need for personalized approaches (Losinski et al., 2015). Additionally, school psychology is influenced by broader social issues such as multicultural training, racial trauma, and anti-Black racism (Rogers, 2006; Proctor, 2022). Understanding and predicting student dropout is vital for enabling educational institutions to intervene early with supports like counselling and targeted academic help, thereby enhancing student success and informing effective retention strategies (Ameri et al., 2016;

Masci et al., 2022; Quadri and Shukor, 2021; Mardolkar and Kumaran, 2020).

Machine Learning and Fairness in School Retention. Traditional approaches to predicting school retention and dropout have largely relied on statistical models, which often fall short in capturing the complex interplay of influencing factors (Westrick et al., 2015; Meeter and M, 2021; Eskreis-Winkler et al., 2014). In contrast, machine learning (ML) offers a data-driven alternative capable of modeling non-linear relationships using techniques such as decision trees, random forests, and boosting algorithms (Chen et al., 2020; Möller et al., 2022). For instance, Dake et al. (2022) showed that random forests achieved a prediction accuracy of 70.98% in identifying student attrition at the University of Education, Winneba. These examples underscore the potential of ML to improve student retention predictions through both simple and ensemble-based methods. However, as ML systems are increasingly deployed in education, their fairness has become a critical concern. Several studies have shown that predictive models may exhibit disparities across demographic groups. Karimi-Haghighi et al. (2021) highlighted differences in error rates when predicting dropout among computer engineering students and proposed the Equalized Generalized False Positive Rate (GFPR) as a bias mitigation strategy. Yu et al. (2020) found that institutional and Learning Management System (LMS) data sources offered strong predictive capabilities, while survey data contributed less, underscoring the importance of equitable data usage. Cock et al. (2023) introduced effective pre-processing techniques for reducing bias in models predicting student success. Building on this foundation, our study focuses on evaluating group fairness (Mehrabi et al., 2021) across several ML classifiers for predicting school dropout. We aim to identify which fairness metrics are most appropriate for this educational context. To the best of our knowledge, this specific intersection of fairness metric evaluation and school retention prediction remains underexplored.

3. Methodology

3.1. Notation

We assume samples (X, A, Y) are drawn i.i.d. from an unknown distribution $\mathcal{D}$. Here: $X \in \mathcal{X} \subseteq \mathbb{R}^d$ denotes non-sensitive features, $A \in \{0, 1\}$ is a binary sensitive attribute, where $A = 1$ indicates the protected group (e.g., female) and $A = 0$ the unprotected group (e.g., male), $Y \in \mathcal{Y} = \{0, 1\}$ is the outcome label, with $Y = 1$ (retention) as the desirable outcome and $Y = 0$ (dropout) as the undesirable outcome. We consider a predictive model $f : \mathcal{X} \times \mathcal{A} \rightarrow \mathcal{Y}$, which maps input features and the sensitive attribute to a predicted label $\hat{Y} = f(X, A)$.

3.2. Fairness Notions

To evaluate the fairness of our machine learning models, we focus on well-known *group fairness* metrics. These notions assess whether model predictions are equitably distributed across demographic groups defined by sensitive attributes (e.g., gender). We adopt these fairness metrics that are widely used in the fair machine learning literature due to their interpretability and broad applicability (Mehrabi et al., 2021; Miron et al., 2020).

Definition 3.1 (Demographic Parity (DP)) *A classifier satisfies demographic parity if the probability of assigning a positive prediction is equal across sensitive groups. Formally,*

$$\mathbb{P}(\hat{Y} = 1 \mid A = 0) = \mathbb{P}(\hat{Y} = 1 \mid A = 1).$$

This definition assumes that the sensitive attribute A (e.g., gender) should not influence the likelihood of receiving a positive outcome. However, demographic parity may lead to unequal false positive rates, which can introduce indirect harm to some groups (Dwork et al., 2012).

Definition 3.2 (Equality of Opportunity (EO)) *A classifier satisfies equality of opportunity if it achieves equal true positive rates across groups, conditioned on the ground truth being positive* Hardt et al. (2016):

$$\mathbb{P}(\hat{Y} = 1 \mid Y = 1, A = 0) = \mathbb{P}(\hat{Y} = 1 \mid Y = 1, A = 1).$$

This metric focuses on ensuring that individuals who truly belong to the positive class (e.g., students who are not at risk of dropout) are treated fairly, regardless of group membership.

Definition 3.3 (Accuracy Parity (Acc)) *A classifier satisfies accuracy parity if its overall accuracy is consistent across groups:*

$$\mathbb{P}(\hat{Y} = Y \mid A = 0) = \mathbb{P}(\hat{Y} = Y \mid A = 1).$$

This reflects whether the model performs equally well (in terms of correct predictions) for all demographic groups Verma and Rubin (2018).

Definition 3.4 (False Positive Rate Parity (FPR)) *A classifier satisfies FPR parity if the probability of predicting a false positive is equal across groups:*

$$\mathbb{P}(\hat{Y} = 1 \mid Y = 0, A = 0) = \mathbb{P}(\hat{Y} = 1 \mid Y = 0, A = 1).$$

This metric is important in contexts where false positives lead to negative consequences.

3.3. Models

All models were implemented using scikit-learn with the following preprocessing pipeline: categorical variables were one-hot encoded, and numerical features were standardized using StandardScaler. Models were trained using default hyperparameters to ensure fair comparison, though we acknowledge that hyperparameter optimization could improve individual model performance. The dataset was stratified during the train-test split to maintain class distribution balance.

3.3.1. LOGISTIC REGRESSION

Logistic regression models the probability of dropout as a function of input features. Given a feature vector $X \in \mathcal{X}$, the predicted probability that a student drops out ($Y = 1$) is:

$$\mathbb{P}(\hat{Y} = 1 \mid X = \mathbf{x_i}) = \sigma(\mathbf{w}^\top \mathbf{x_i} + b), \quad \text{where } \sigma(z) = \frac{1}{1 + e^{-z}}.$$

The parameters $\mathbf{w}$ (weights) and b (bias) are optimized via maximum likelihood estimation:

$$\log L(\mathbf{w}, b) = \sum_{i=1}^{n} \left[y_i \log \mathbb{P}(\hat{Y}_i = 1 \mid X = \mathbf{x_i}) + (1 - y_i) \log \left(1 - \mathbb{P}(\hat{Y}_i = 1 \mid X = \mathbf{x_i}) \right) \right].$$

Each coefficient in $\mathbf{w}$ represents the contribution of its corresponding feature to the likelihood of dropout, providing interpretability into which factors most influence student retention.

3.3.2. DECISION TREES

A decision tree predicts the target class by learning a sequence of decision rules inferred from training data. The model starts at a root node and recursively splits the data based on selected features until a stopping criterion, such as maximum depth or minimum leaf size is met (Tan et al., 2006). Our work uses the *Iterative Dichotomiser 3 (ID3) algorithm*, which builds the tree using a top-down greedy search that selects the best feature at each step according to a splitting criterion.

$$\text{Entropy}(S) = - \sum_{i=1}^{k} p_i \log_2(p_i), \tag{1}$$

where p_i is the proportion of samples in class i within the subset S. Alternatively, the Gini index is another commonly used impurity measure:

$$\text{Gini}(S) = 1 - \sum_{i=1}^{k} p_i^2, \tag{2}$$

where lower values indicate purer node splits (Ichihashi et al., 1996). The final prediction $\hat{Y}$ is obtained by traversing the tree according to the feature values in $X \in \mathcal{X}$ until a leaf node is reached.

3.3.3. RANDOM FOREST

Random Forest is an ensemble learning method that combines the outputs of multiple decision trees to improve prediction accuracy and robustness. Each tree $f_i : \mathcal{X} \to \hat{\mathcal{Y}}$ is trained on a bootstrapped sample of the training data and a random subset of features, promoting variance reduction and preventing overfitting. Given an input instance $X \in \mathcal{X}$, the final prediction $\hat{Y}$ is obtained through majority voting across k decision trees:

$$\hat{Y} = \text{mode}\left(\{f_1(X), f_2(X), \ldots, f_k(X)\}\right),$$

where $f_i(X) \in \{0, 1\}$ denotes the prediction of the i-th tree. This ensemble approach effectively captures complex patterns in the data, making it particularly suitable for heterogeneous features such as academic history and socio-economic indicators.
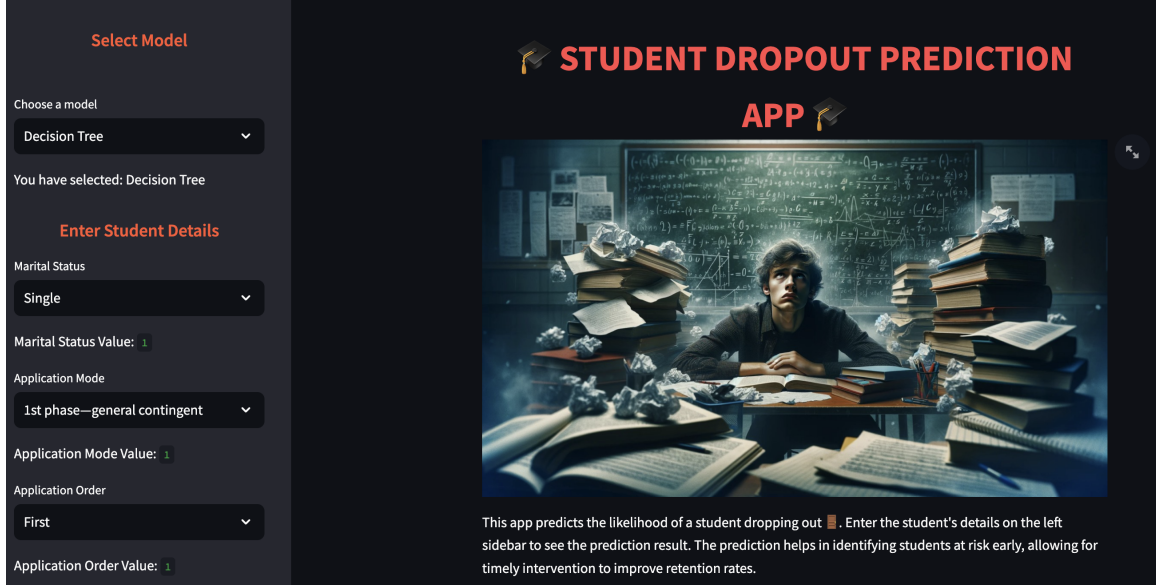
Figure 1: Frontend of the school dropout classifier. The classifier is available at https://abubakari-fair-student-retention.hf.space/.

3.4. Front-End Web Application

To enable interactive and accessible student dropout forecasting, we developed a web-based application that integrates machine learning predictions with real-time user input. The system is designed for high usability and deployability. Trained models are serialized using `joblib` to enable efficient loading during inference. Dependencies are managed via a `requirements.txt` file to ensure environment consistency. The application is built with Streamlit, providing a user-friendly interface for inputting student data and receiving instant dropout predictions. It is deployed on Hugging Face Spaces and integrated with GitHub for version control and continuous updates. Figure 1 shows the interface. In addition to binary classification, the system flags predictions as "fair" or "unfair" using internal fairness metrics, thereby improving transparency and trust in model outputs. This design enables real-time, accessible insights into student dropout likelihood while incorporating fairness assessments for ethical decision support.

4. Results

4.1. Dataset Overview

We use a dataset comprising $4,424$ undergraduate student records from academic years 2008/2009 to 2018/2019, spanning various disciplines (Realinho et al., 2022). The dataset includes 23 complete features capturing demographic, socioeconomic, and academic factors. Key variables include marital status, course enrollment, attendance type, nationality, special education needs, debt status, scholarship status, and first-semester performance, making it well-suited for analyzing predictors of student dropout.

4.2. Exploratory Insights and Model Evaluation

Exploratory data analysis (EDA) revealed that most students were enrolled full-time. Dropout rates varied across courses and nationalities, while financial indicators such as debt and scholarships were strongly associated with dropout risk. These insights informed our feature selection and preprocessing pipeline. We evaluated logistic regression (LR), decision trees (DT), and random forests (RF) using classification metrics (accuracy, F1 score, recall, precision) and group fairness metrics (demographic parity, equality of opportunity, false positive rate) with Table 1 summarizing the model performance [1]

Table 1: Model performance evaluation. The models include (LR: Logistic Regression, RF: Random Forest, and DT: Decision Trees). Performance comparison by column.

Model	Acc	F1	Recall	Precision
Logistic Regression	**0.883**	**0.880**	**0.883**	**0.882**
Random Forest	0.878	0.874	0.878	0.877
Decision Tree	0.816	0.816	0.816	0.816

Table 1, shows that the Logistic Regression model outperforms the others across all evaluation and fairness metrics. It achieves the highest accuracy (88.3%), F1 score (88.0%), recall (88.3%), and precision (88.2%). While random forest yields slightly lower values across the board, it remains competitive. Decision trees, however, demonstrate notably lower performance, suggesting they are less suitable for this prediction task.

	LR		RF		DT	
Metric	Male	Female	Male	Female	Male	Female
DP (↑ better)	0.188	**0.390**	0.185	**0.399**	0.236	**0.463**
EO(↑ better)	0.721	**0.732**	0.706	**0.737**	0.677	0.737
Acc(↑ better)	**0.912**	0.825	**0.909**	0.821	**0.844**	0.757
FPR(↓ better)	0.032	**0.093**	**0.032**	0.106	0.106	**0.225**

Table 2: Comparison of models within the different sensitive groups.

Table 2 shows subgroup results across gender. LR and RF models exhibit moderate disparities, while DT reveals substantial fairness gaps, particularly in demographic parity (DP) and false positive rate (FPR). This suggests DT may propagate more bias compared to LR and RF. Table 3 highlights gender disparities across models. Decision Trees (DT) show the highest disparities in demographic parity (0.2270), equality of opportunity (0.0600), and false positive rate (0.1182), indicating significant bias. In contrast, Logistic Regression (LR) achieves the lowest disparities in EO (0.0103) and matches Random Forest (RF) in FPR (0.0612), suggesting more equitable performance. All models favor male predictions in accuracy, with RF showing the largest gap (-0.0890). Overall, LR is the most balanced model, while DT raises the greatest fairness concerns.

---

1. Code is available at: https://github.com/KANUBALAD/ML-Fairness-in-School-DropOut-

Table 3: Differences in Evaluation Metrics between Male and Female for Different Models using results from Table 2. The accuracy is showing negative because we took the difference without absolute values. We compare row by row for the different fairness metrics, and the lower the score,s the better.

Metrics(Difference)	LR	RF	DT
DP(↓ better)	0.2017	**0.1995**	0.2270
EO(↓ better)	**0.0103**	0.0156	0.0600
Acc(↓ better)	**-0.0867**	-0.0890	-0.0870
FPR(↓ better)	**0.0612**	**0.0612**	0.1182

★ LR not only provides the best predictive performance but also the most equitable outcomes across gender, making it the most suitable model for fair and accurate dropout prediction in this context.

5. Discussion

Trade-Offs Between Predictive Accuracy and Fairness. While Logistic Regression (LR) achieved the highest predictive accuracy among the models evaluated, our findings underscore that accuracy alone is not a sufficient criterion for responsible model deployment. The relatively low disparities in fairness metrics such as Equality of Opportunity (EO) and accuracy difference suggest that LR may offer a reasonable balance between performance and equity. However, its advantages must be contextualized within the broader goals of educational equity and institutional accountability.

Disparities and Their Implications. The disparities observed across models have concrete implications for educational equity. Decision Trees' 22.7% demographic parity gap means that female students are substantially more likely to be flagged as at-risk regardless of their actual risk level. In practical terms, if 1000 students were evaluated, this could result in approximately 227 more female students being unnecessarily targeted for interventions, potentially leading to stigmatization and inefficient resource allocation. Conversely, the model's 11.82% false positive rate disparity could result in missed opportunities to support genuinely at-risk male students. These are not merely statistical artefacts but represent real potential for educational harm.

Ethical Considerations in Educational Contexts. Educational institutions operate in domains where algorithmic decisions can have long-term impacts on students' lives. Models that over-identify risk in one group may lead to stigmatization or unnecessary intervention, while under-identifying risk in another may result in missed opportunities for support. These ethical risks call for ongoing audits, stakeholder engagement, and transparency in how predictions are used to inform decisions.

Institutional and Policy Recommendations. Institutions should adopt models that balance accuracy with fairness, integrating fairness metrics into evaluation pipelines. Training stakeholders on ML interpretation and establishing accountability mechanisms are essential steps toward equitable and responsible deployment. Our findings show that

fairness-aware models like LR, supported by institutional safeguards, can promote more inclusive student success strategies.

5.1. Fairness Metric Suitability for Education

Our analysis suggests that different fairness metrics serve distinct purposes in educational contexts. Equality of Opportunity emerges as particularly relevant because it ensures that students who are truly not at risk (Y=1) are equally likely to be correctly identified across gender groups, preventing systematic under-support of any demographic. False Positive Rate parity is crucial because incorrectly flagging students as at-risk can lead to unnecessary interventions and potential stigmatization. Demographic Parity, while conceptually appealing, may be less suitable for education as it ignores base rate differences in dropout risk that may legitimately vary across groups due to external socioeconomic factors. Accuracy Parity provides an overall fairness assessment but may mask specific types of discrimination. Based on our findings, we recommend prioritizing Equality of Opportunity and False Positive Rate parity for educational applications, with Logistic Regression showing the best balance across these critical metrics

6. Conclusion

Our findings have immediate practical implications for educational institutions implementing ML-based early warning systems. Institutions should prioritize Logistic Regression models while implementing continuous fairness monitoring using Equality of Opportunity and False Positive Rate metrics. The relatively small fairness gaps observed suggest that with proper model selection and ongoing evaluation, it is possible to achieve both high predictive accuracy and equitable outcomes in student retention applications.

References

Muhib Al-kmali, Hamzah Mugahed, Wadii Boulila, Mohammed Al-Sarem, and Anmar Abuhamdah. A machine-learning based approach to support academic decision-making at higher educational institutions. In *2020 international symposium on networks, computers and communications (ISNCC)*, pages 1–5. IEEE, 2020.

Sattar Ameri, Mahtab Jahanbani Fard, Ratna Babu Chinnam, and Chandan K. Reddy. Survival analysis based framework for early prediction of student dropouts. 2016. doi: 10.1145/2983323.2983351.

Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. In *Ethics of data and analytics*. Auerbach Publications, 2022.

Lovenoor Aulck, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. Predicting student dropout in higher education. *arXiv preprint arXiv:1606.06364*, 2016.

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

Jiale Chen, Jason Hartline, and Onno Zoeter. Fair grading algorithms for randomized exams. *arXiv preprint arXiv:2304.06254*, 2023.

Lijia Chen, Pingping Chen, and Zhijian Lin. Artificial intelligence in education: A review. *Ieee Access*, 2020. doi: 10.1109/access.2020.2988510.

Jade Maï Cock, Muhammad Bilal, Richard Davis, Mirko Marras, and Tanja Kaser. Protected attributes tell us who, behavior tells us how: A comparison of demographic and behavioral oversampling for fair student success modeling. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, pages 488–498, 2023.

Delali Kwasi Dake, Charles Buabeng-Andoh, et al. Using machine learning techniques to predict learner drop-out rate in higher educational institutions. *Mobile Information Systems*, 2022, 2022.

Jeffrey Dastin. Amazon scraps secret ai recruiting tool that showed bias against women. In *Ethics of data and analytics*, pages 296–299. Auerbach Publications, 2022.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.

Lauren Eskreis-Winkler, Elizabeth P. Shulman, Scott A. Beal, and Angela Duckworth. The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, 2014. doi: 10.3389/fpsyg.2014.00036.

Barry A. Friedman and Rhonda G. Mandel. Motivation predictors of college student academic performance and retention. *Journal of College Student Retention Research Theory & Practice*, 2011. doi: 10.2190/cs.13.1.a.

Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29, 2016.

Brody Heritage, Christopher Ladeira, and Andrea R. Steele. The development and pilot of the university student embeddedness (use) scale for student retention within universities: Validation with an australian student sample. *Higher Education*, 2022. doi: 10.1007/s10734-022-00813-z.

Hidetomo Ichihashi, T Shirai, Kazunori Nagasaka, and Tetsuya Miyoshi. Neuro-fuzzy id3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy sets and systems*, 81(1):157–167, 1996.

Stephen Kampf and Eric J. Teske. Collegiate recreation participation and retention. *Recreational Sports Journal*, 2013. doi: 10.1123/rsj.37.2.85.

Marzieh Karimi-Haghighi, Carlos Castillo, Davinia Hernandez-Leo, and Veronica Moreno Oliver. Predicting early dropout: Calibration and algorithmic fairness considerations. *arXiv preprint arXiv:2103.09068*, 2021.

Mickey Losinski, Nicole Wiseman, Sherry A. White, and Felicity Balluch. A meta-analysis of video-modeling based interventions for reduction of challenging behaviors for students with ebd. *The Journal of Special Education*, 2015. doi: 10.1177/0022466915602493.

Mahesh Mardolkar and N. Kumaran. Forecasting and avoiding student dropout using the k-nearest neighbor approach. *Sn Computer Science*, 2020. doi: 10.1007/s42979-020-0102-0.

Chiara Masci, Mirko Giovio, and Paola Mussida. Survival models for predicting student dropout at university across time. 2022. doi: 10.36315/2022v1end043.

Martijn Meeter and van Brederode M. Predicting retention in higher education from high-stakes exams or school gpa. 2021. doi: 10.31234/osf.io/h537g.

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54 (6):1–35, 2021.

Thomas Miconi. The impossibility of" fairness": a generalized impossibility result for decisions. *arXiv preprint arXiv:1707.01195*, 2017.

Marius Miron, Songül Tolan, Emilia Gómez, and Carlos Castillo. Addressing multiple metrics of group fairness in data-driven decision making. *arXiv preprint arXiv:2003.04794*, 2020.

Annette Möller, Ann Cathrice George, and Jürgen Groß. Predicting school transition rates in austria with classification trees. *International Journal of Research & Method in Education*, 2022. doi: 10.1080/1743727x.2022.2128744.

Yuri Nieto, Vicente García-Díaz, Carlos Montenegro, and Rubén González Crespo. Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Computing*, 23:4145–4153, 2019.

Yuri Vanessa Nieto Acevedo, Vicente García Díaz, Carlos Enrique Montenegro, et al. Decision-making model at higher educational institutions based on machine learning. *Journal Of Universal Computer Science, 25*, 2019.

Juan C Perdomo, Tolani Britton, Moritz Hardt, and Rediet Abebe. Difficult lessons on social prediction from wisconsin public schools. *arXiv preprint arXiv:2304.06205*, 2023.

Sherrie L. Proctor. From beckham until now: Recruiting, retaining, and including black people and black thought in school psychology. *School Psychology International*, 2022. doi: 10.1177/01430343211066016.

Adedayo Taofeek Quadri and Nurbiha A. Shukor. The benefits of learning analytics to higher education institutions: A scoping review. *International Journal of Emerging Technologies in Learning (Ijet)*, 2021. doi: 10.3991/ijet.v16i23.27471.

Valentim Realinho, Jorge Machado, Luís Baptista, and Mónica V Martins. Predicting student dropout and academic success. *Data*, 7(11):146, 2022.

Steven B. Robbins, Kristy J. Lauver, Huy Le, Daniel Cochece Davis, Ronelle Langley, and Aaron H. Carlstrom. Do psychosocial and study skill factors predict college outcomes? a meta-analysis. *Psychological Bulletin*, 2004. doi: 10.1037/0033-2909.130.2.261.

Margaret R. Rogers. Exemplary multicultural training in school psychology programs. *Cultural Diversity and Ethnic Minority Psychology*, 2006. doi: 10.1037/1099-9809.12.1. 115.

Latanya Sweeney. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. *Queue*, 11(3):10–29, 2013.

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Classification: basic concepts, decision trees, and model evaluation. *Introduction to data mining*, 1:145–205, 2006.

Sahil Verma and Julia Rubin. Fairness definitions explained. In *Proceedings of the international workshop on software fairness*, pages 1–7, 2018.

Paul Westrick, Huy Le, Steven B. Robbins, Justine Radunzel, and Frank L. Schmidt. College performance and retention: A meta-analysis of the predictive validities of act¡sup¿Ⓡ¡/sup¿scores, high school grades, and ses. *Educational Assessment*, 2015. doi: 10.1080/10627197.2015.997614.

Renzhe Yu, Qiujie Li, Christian Fischer, Shayan Doroudi, and Di Xu. Towards accurate and fair prediction of college success: Evaluating different sources of student data. *International educational data mining society*, 2020.