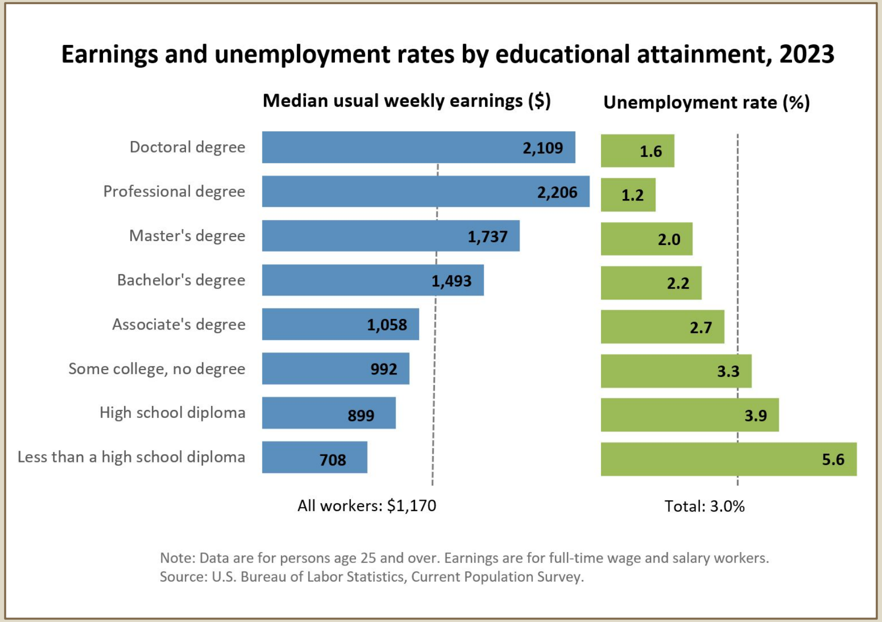
Predicting Dropout Rates: A Study of Equity and Bias

By Ananya Kulshrestha and Ronak Saluja

"Once you've completed your education, you will have the foundation you need to build a successful life"

- Former First Lady, Michelle Obama

The Importance of Education



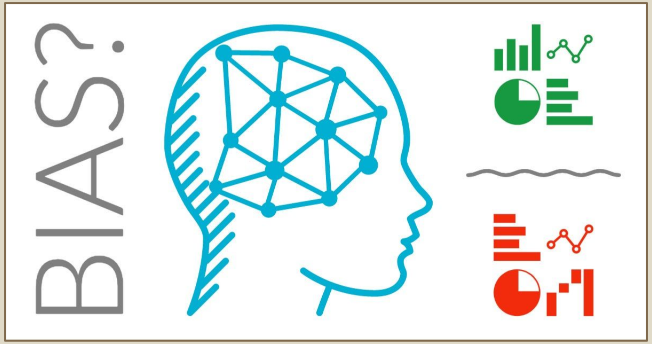
- **Educational level** directly impacts future **financial and personal well-being**
- Exacerbated by **social inequalities**
- Vital to detect students **at risk of dropping out**

Pitfalls of Existing Predictive Models

- Predict failure more often for Black & Hispanic students
- Overestimate success for White & Asian students
- Students with “socially relevant features” disadvantaged
- ML research prioritizes performance over fairness

“We took bad data in the first place, and then we used tools to make it worse … [i]t's just been a self-reinforcing loop over and over again.”

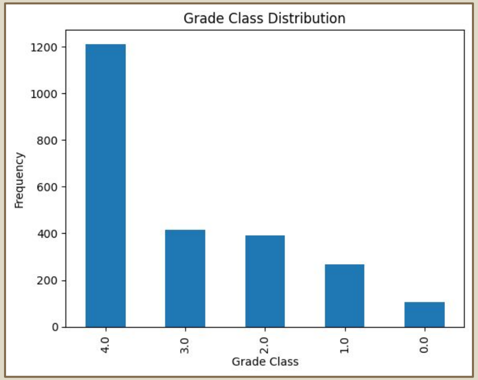
- Katy Weathington, algorithmic bias researcher



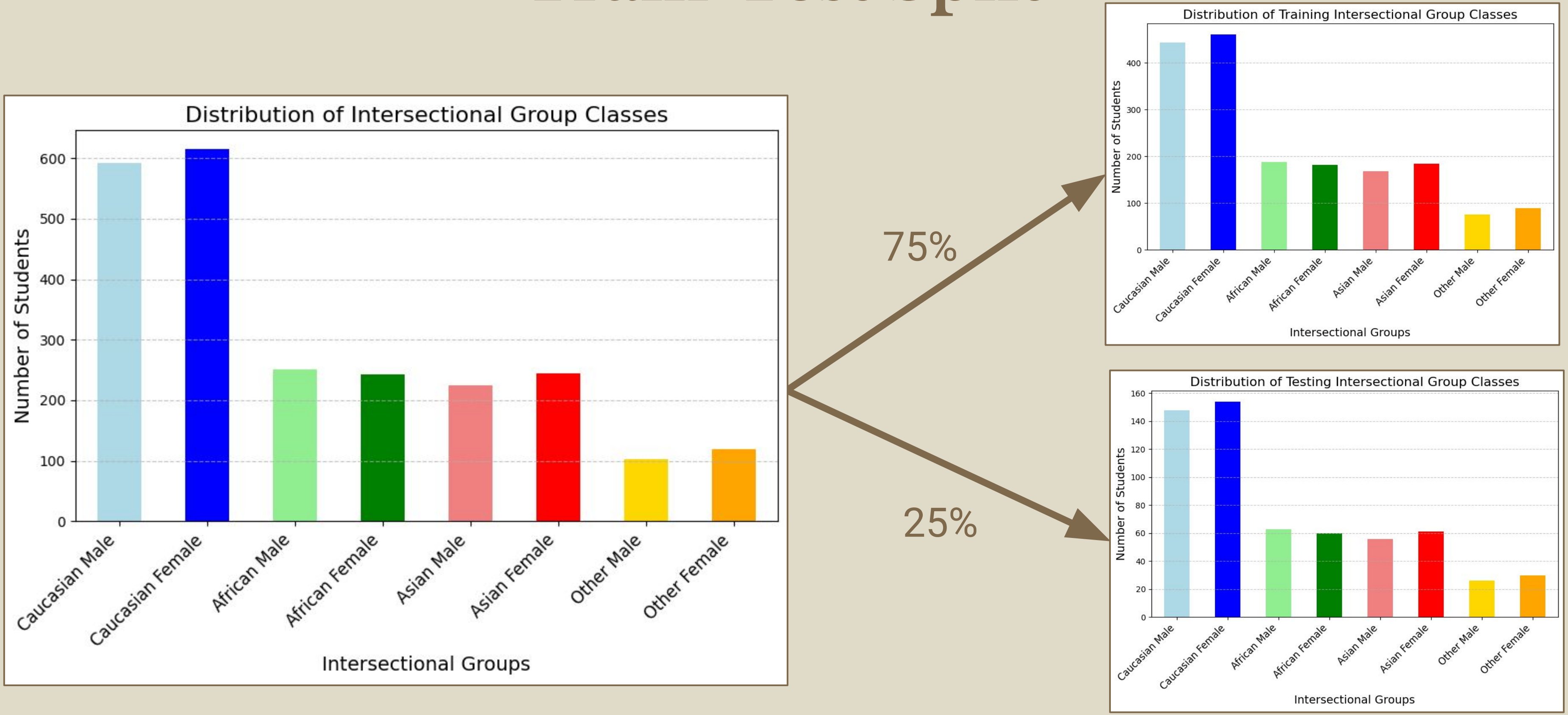
How are machine learning models that are used to predict academic performance perpetuate or mitigate biases based on race and gender? Does excluding or including “socially relevant” features affect model accuracy and fairness?

Dataset and Preprocessing

- **Kaggle Dataset: Students Performance**
  - About 2.5K students' data of ages between 15 and 18
  - Includes demographics, parental involvement, academic performance, etc.
- **Preprocessing**
  - Binary classification: At dropout risk or not
  - Used GradeClass score of 4 (i.e. GPA < 2.0) as signifying dropout risk
    - Scores 0, 1, 2, 3 were grouped
- **Features:**
  - Focused on effects of Gender and Ethnicity
  - Removed "GPA" features



Train-Test Split

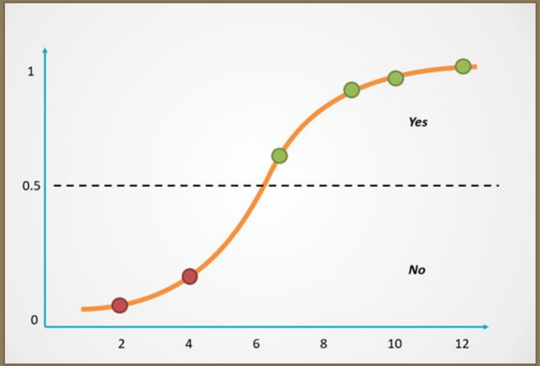


Experiments and Models

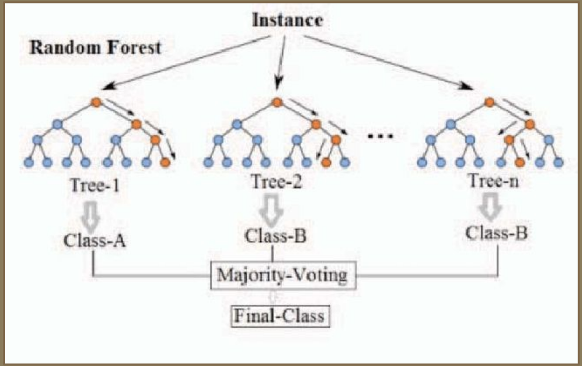
Trained **three models** of varying complexity **four times each**:

1. **All** features
2. Removed the **Gender** Feature
3. Removed the **Ethnicity** Feature
4. Removed both the **Ethnicity and Gender** Features

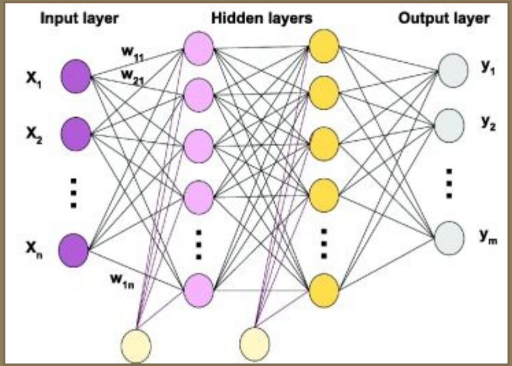
Logistic Regression



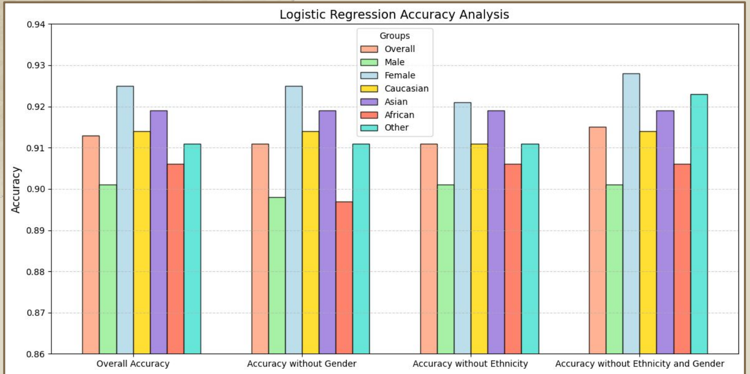
Random Forest

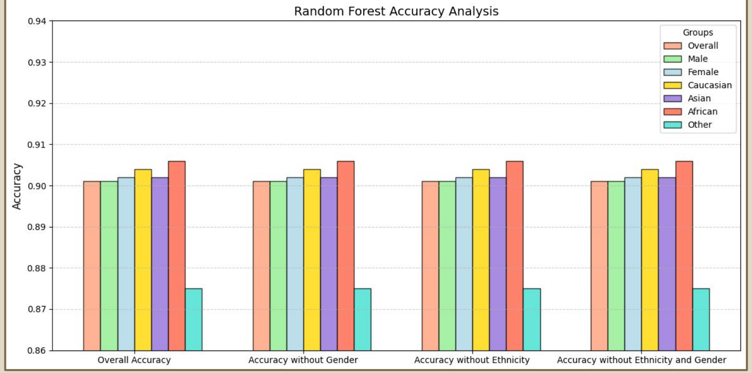


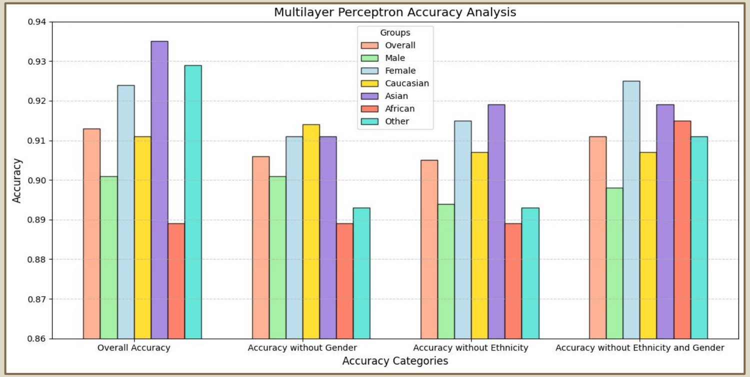
Multilayer Perceptron



Results



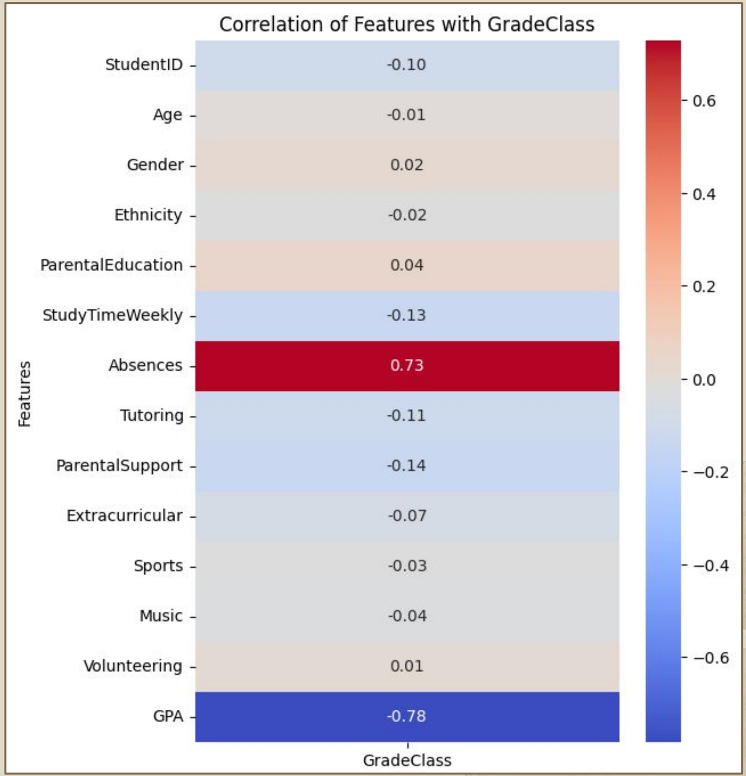




High, consistent performance across all models and demographics

The Dominance of Absences

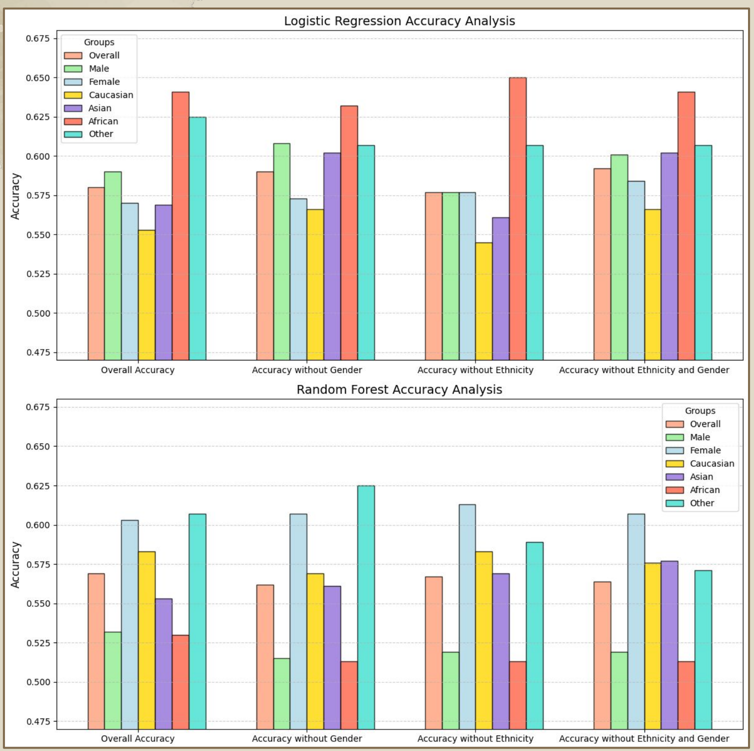
The "Absences" feature is ~<u>80 - 90%</u> of a trained model's feature importance

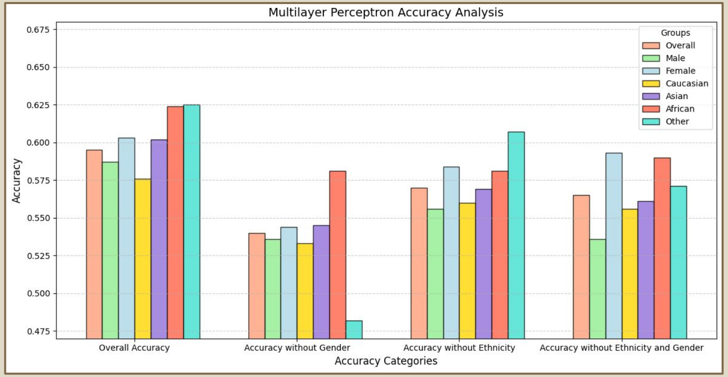


Will "socially relevant" features play more of a role if this feature is removed?

Will performance still persist without this feature?

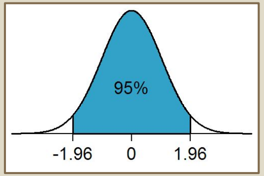
Removal of Absences



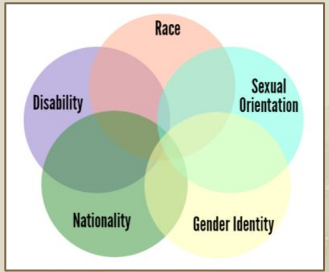


Significant dip in accuracy
(From ~85-95% to ~50-65%)

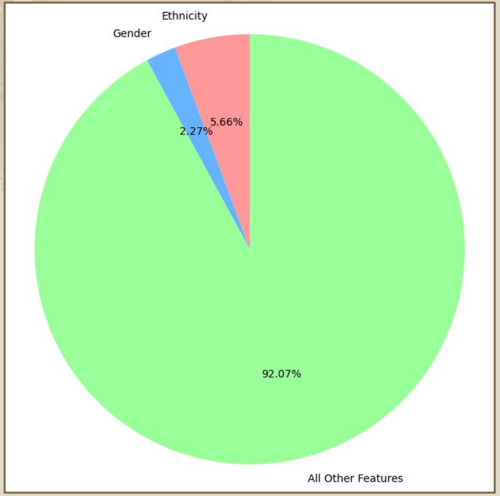
Statistically Significance Differences

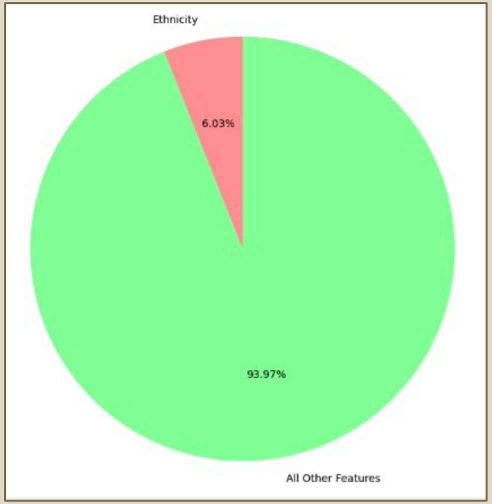


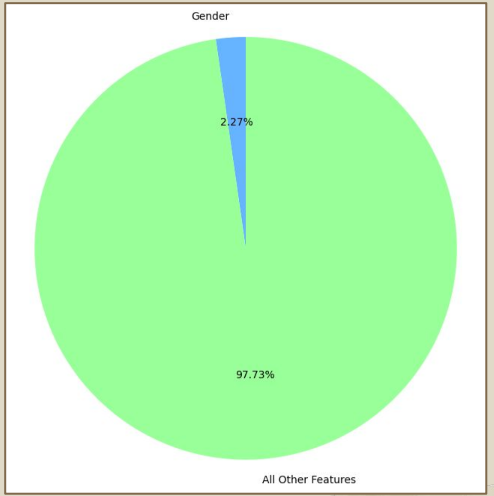
- For each model, conducted **Pairwise Z-Test** on two axes:
  - Model configuration: with different sets of features
  - Demographic groups e.g. African and Caucasian
- Key Finding:
  - Only one MLP model show significance
    - “Other” ethnicity: All features vs. Excluding Gender
- Societal Implication:
  - Highlights potential for **intersectional bias**
  - Gender difference may affect underrepresented groups more than majority groups



Feature Importances in RF Models







"Socially relevant" features play a **small role** in influencing results

Final Takeaways

1) A model is only as good as its data

2) High performance and fairness must coexist

3) Small decisions matter: design AI ethically

Works Cited

https://www.bls.gov/emp/chart-unemployment-earnings-education.htm

https://obamawhitehouse.archives.gov/photos-and-video/video/2013/11/12/first-lady-michelle-obama-speaks-power-education

https://www.aera.net/Newsroom/Study-Algorithms-Used-by-Universities-to-Predict-Student-Success-May-Be-Racially-Biased

https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/?utm_source=chatgpt.com

https://arxiv.org/pdf/2106.15590

Dataset: 10.34740/kaggle/ds/5195702

Thank you

We’re open to taking any questions now!