

### **Stage 3: Initial Results & Challenges**

Joya Wheatfall-Melvin

PPOL 5204: Data Science II

Professor Ioannis Ziogas

March 23, 2025

**Research Design & Methodology:** Discuss your analytical plan.

*Alternative Schools SCH\_STATUS\_ALT:* is a public elementary or secondary school that addresses the needs of students that typically cannot be met in a regular school program and is designed to meet the needs of students with academic difficulties, students with discipline problems, or both students with academic difficulties and discipline problems. Alternative education schools may be sited in locations other than a traditional school building such as hospitals, mental health centers, jails, or juvenile detention centers.

*Justice (JUST)* A justice facility is a public or private facility that confines pre-adjudicated/pre-convicted individuals, postadjudicated/post-convicted individuals, or both. Justice facilities include short-term (90 calendar days or less) and long-term (more than 90 calendar days) facilities, such as correctional facilities, detention centers, jails, and prisons. These facilities may confine juveniles (individuals typically under 21 years of age), adults (individuals typically 21 years of age and older), or both. Some states and jurisdictions include individuals younger than age 21 as adults due to statute/legislation and/or justice procedures. For the CRDC, only individuals up to 21 years of age who are confined in justice facilities are of interest.

- While alternative schools serve as a transitional period in one's educational life (some students stay for days, others for years), it still has a significant effect on one's educational attainment and/or educational perspective. The U.S. Accountability Office finds that “certain groups of students attend alternative schools in greater proportions than they do other schools”, specifically black girls and boys.

Specification of Research Question: **“How do school-provided opportunities and climate conditions shape student outcomes for Black girls in alternative schools?”**

-Still speaks to the integrity of alternative schooling for black girls.

**\*After Modelling, I have to modify my Target and Feature Variables\***

**Measuring “Integrity” (Outcomes):** Negative outcomes that Black Girls experience

- Alternative School Graduation Rates (Not Applicable)
  - Non-intersectional data struggles: Most local database demographics were either divided by Race or Gender
  - “Alternative high schools have substantially lower 4-year graduation rates” (GAO-20-310, 2020)
- Low Retention Rates
  - Aggregated all Retention (K-12) into rates per school
- Chronic Absenteeism
- Expulsions
- Suspensions
- Arrests

- Corporal Punishment
- Seclusion

**Measuring Feature Variables (Access):** Reflecting access, inclusion, and program structure

- College Career Readiness Indicators
- Program Access
  - Gifted
  - Accelerated Classes
  - ACT/SAT Prep
  - Dual Enrollment
- Add proportions of the black female population in these courses as well
  - Course Involvement

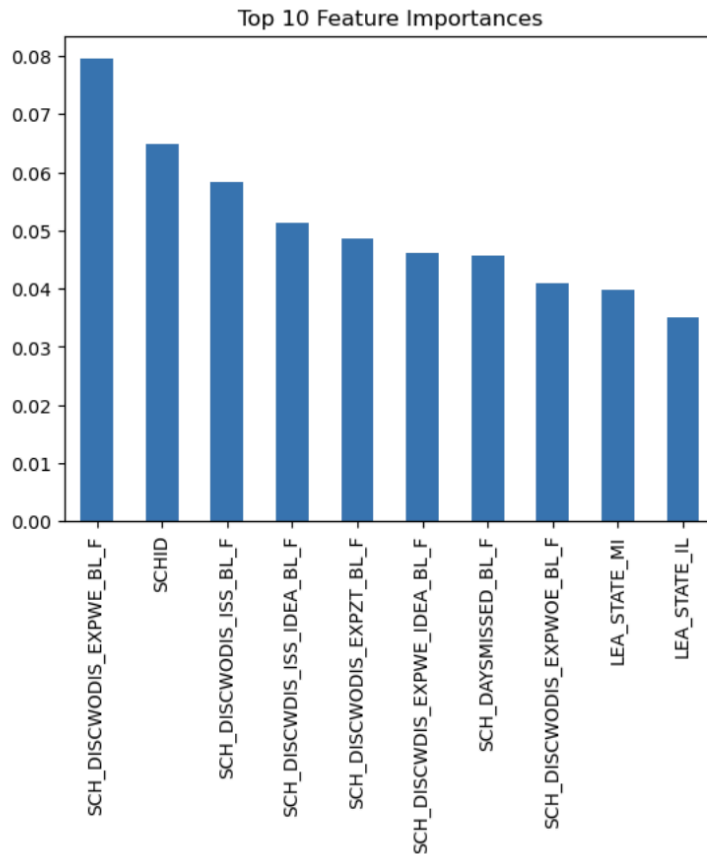
This would be a classification problem: We want to predict which school characteristics are better suited to pushing black female students to thrive in alternative schools. (Since they are the overwhelming majority)

- **Random Forests**
  - Information-based classifier
  - Most necessary in answering *which* factors most influence retention for black female students
- **Logistic Regression**
  - Can interpret the data by showing *how* certain factors influence retention for black female students
  - However, could be sensitive to multicollinearity (I believe a lot of my features may be correlated) ‘

**Results:** This section should include your initial results. Specifically:

- Random Forest Output:
  - The model is perfect for classifying the majority class (no retention)
    - Solely because the classes are imbalanced.
  - Identified 3 out of 5 true retention (60% recall)
  - Precision = 1.00, meaning every time it predicted retention, it was correct
    - Class imbalance
  - F1 score = 0.75
  - Accuracy = 1.00
    - Class imbalance; meaning it predicted “no retention” every time

- Interpretation: Analyze and interpret your results and provide any relevant graphs.



- SCHID is ranked the highest; this is an effect of not accurately pre-processing.
- Discipline and Punishment-related features rank high as well due to having quality data.
  - The target variable is Retention which is highly correlated with Punishment variables.
  - This will change once I rearrange my target and feature variable through more preprocessing.

**Limitations / Challenges:** Address the limitations of your models and any challenges you have encountered. In particular:

- This is a conflicting output; it doesn't make sense and here's why:
  - I need to continue preprocessing this data; there are a lot of missing values due to reserve codes, and demographic population. I believe preprocessing and standardizing this data will take me the longest, but will ensure that I have a good model.
  - This model is using "Retention" (Grades 9-12) only. In order to have a full data set, I need to include all of the other harmful indicators for a full target variable.

- Because of this, the target variable was extremely unbalanced, even with the threshold I introduced. Retention cannot be my only target variable.
- Figuring out how to resolve missing data will be a struggle as well. I do not want to impute a mass majority of the data because of bias.
  - Hence, including more indicators as the target variable to bulk up data validity.

## References

- Anderson, M. D. (2024, August 13). *How Black Girls Are Harmed in Schools*. Education Writers Association. Retrieved February 2, 2025, from <https://ewa.org/data-research-tips/how-black-girls-are-harmed-in-schools>
- Cabral, B., Annamma, S. A., & Morgan, J. (2023). “When You Carry a Lot”: The Forgotten Spaces of Youth Prison Schooling for Incarcerated Disabled Girls of Color. *Teachers College Record*, 125(5), 95-113. <https://doi.org/10.1177/01614681231181816>
- Hartocollis, A. (2024, December 16). *Black Student Enrollment at Harvard Law Drops by More Than Half*. The New York Times. Retrieved February 2, 2025, from <https://www.nytimes.com/2024/12/16/us/harvard-law-black-students-enrollment-decline.html>
- National Center for Education Statistics. (n.d.). *Fast Facts: Degrees conferred by race/ethnicity and sex (72)*. National Center for Education Statistics. Retrieved February 2, 2025, from <https://nces.ed.gov/fastfacts/display.asp?id=72>
- U.S. Department of Education. (n.d.). *Civil Rights Data Collection Office for Civil Rights*. Civil Rights Data | U.S. Retrieved February 2, 2025, from <https://civilrightsdata.ed.gov/>