

James Yang

DATA 512

11/19/22

Activity 8

Step 0:

James Yang – Facilitator, timekeeper, notetaker, and submitter.

Step 1:

This product is for colleges, school analytics, and people working in the education sector of the world. The primary stakeholders would be schools and administration, and the secondary stakeholders would be the students themselves having a stake in how their performance is being evaluated.

The purpose of this system is solving the discrepancy between GPA and how a student truly performs. One could take all general education classes and be considered a 4.0 student whereas another student could be taking all AP courses and have a 4.0 as well. Which student is distinguished through the GPA? Neither. This product is intended to differentiate these students and put them in separable classes for colleges, professors, and other clients to analyze. There is more separation between students, given on a regularized scale to allow for more distinguishability.

The contextual factors on how people will experience the product will vary. Stakeholders may begin to manipulate how their systems react to such a product. The stakeholders that will directly interact with the product are the school administration and teachers because they are the ones who assign the students grades which are indicative of a student's overall performance.

The students will indirectly interact with the product, creating training data for the product and their performance may impact the coming future performance of generations. This may be due to change in expectation, similar features, etc.

Additional details about the project:

- Identifying a set of training data that has the least amount of bias.
- Creating a logistic regression of the expectations could allow for prediction of how a student is going to perform with similar characteristics.
- Identifying weights of features of students
- Retraining models based on higher residual expectations.

- Finding scalable features that could be identified for students such as areas of interests.
- Testing data should be resampled consistently based on schools.

Step 2:

There are many ethical risks that could be imposed on this product. The first ethical imbalance is misrepresentation of a general population based on inherent bias. School is a place of development, which means their features have increased variance. That means that if a student with similar features has a trajectory of lower performance, it is not fair to assume that the student won't develop because of his featural past.

Checklist:

| Checklist Question | Example of Ethical Issues |
|--|---|
| Collection bias: Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those? | This system may fail to collect students of all samples, given that students are prone to change their behavior throughout school. Having a small sample size of students is a difficult representation for a whole. |
| Data Security: Is the data going to be encrypted in some manner that will be difficult to obtain or interpret from an outside source? | This system may reveal true student data to the world and impact their performance and reputation in the future. |
| Data Retention: Is there a schedule or plan to delete the data after it is no longer needed? | This system may provide information that is dated and not needed to be retained for the public to know for longer durations of time. For example, a famous student who graduated 20 years ago should not have his information on file for interpretation to the public. |
| Honest Representation: Are the statistics and reports representing honest student performance? | A lot of students and teachers may begin to skew the data for their performance in an effort to bring them to a better school. This in turn will result in a very unfortunate circumstance of altering their performance rather than their true learning of a subject. |
| Explainability: Is the model understandable and do the decisions make sense? | A model may put a student on lower scale but if it does not justify its reasoning, then there is no reason to believe its value and then the entire product becomes more obsolete. |

Step 3:

I have chosen to use the Google guidebook. The first pattern that they have is **determining if AI adds value**. This definitely provides a guideline for explainability, collection bias, and honest representation. We would implement the product based on their idea of a heuristic-based solution with maintenance of predictability. The product should definitely describe step by step why a

student is getting the score that they are receiving and why they are located in the rank that they are.

The second pattern that they have is **being transparent about privacy and data settings**.

Students will want to know what they are doing wrong and why their score is being translated in the way they are. Having students understand the AI behind their performance might impact how they believe the system and how much they trust that it is driving them in the right direction.

The last pattern that they have is **go beyond in the moment explanations**. There needs to be deeper explanations than just the rank of the student. They need to understand how they can improve and the explanations behind their score. This adds to the data security aspect as well because they need to know that their score is private and hidden from other data sources. The model needs to reassure the user that everything is safe and that your score is confidential.