Project Proposal Draft

Heilmeier Question 1: What are you trying to do? Articulate your objectives using absolutely no jargon.

We are investigating the impact of social and economic factors on academic performance in New York state. We feel that the state of the art is somewhat lacking and that we can provide a better understanding of these factors through applied machine learning and thoughtful interactive data visualization.

Heilmeier Question 2: How is it done today; what are the limits of current practice?

Per [Okioga, Farooq, Hearn] the current models are statistics-based, with ANOVA and T-Test applied to academic performance output.

Heilmeier Question 3: What's new in your approach? Why will it be successful?

According to

Heilmeier Question 4: Who cares?

<Anyone have a paper that talks about impact of SES Studies>

Heilmeier Question 5: If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)?

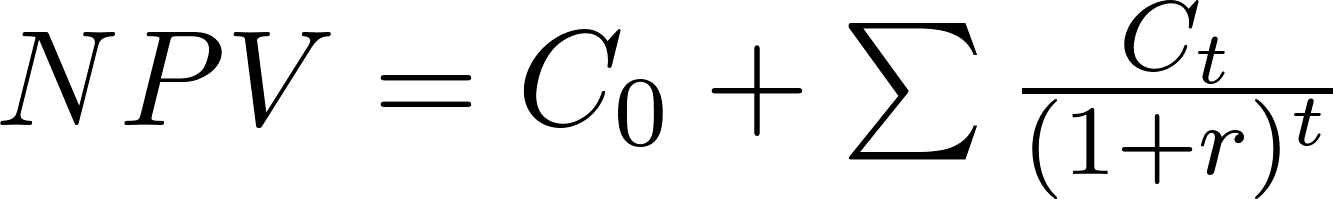
Based on [insert paper name here] the impact of SES studies on academic performance is <blah>

Heilmeier Question 5: What are the risks and payoffs?

ML Models that associate SES and academic performance are often biased and racist, not taking into account confounding variables - reference 0.5 article on racist AI. Black box risk.

Heilmeier Question 6: How much will it cost?

Based on How to Get a Free Lunch: A Simple Cost Model for Machine Learning Applications the cost to data ratio is 0.

[](https://www.codecogs.com/eqnedit.php?latex=NPV%20%3D%20C_0%20%2B%20%5Csum%7B%5Cfrac%7BC_t%7D%7B(1%2Br)%5Et%7D%7D#0)

NPV is the net present value

C0 is the initial cost of equipment, since we will be working with systems we already own and data is freely available, there is no initial cost.

Ct is the decision cost, which is also zero since we won’t be making changes to the system or data we are using

Heilmeier Question 7: How long will it take?

From Lessons from My First Two Years of AI Research, time will depend on many factors. This is a breakdown of a sample timeline:

Reading/Research:

2 weeks - reading and understanding past research will equip us with a better understanding of the subject. Understanding past research shortcoming and limitations can help us find novel approaches.

Conversation/Videos/Conference talks

1 week - Any knowledge gaps should be supplemented by having a conversation or other forms of interactions with a subject matter expert.

Data Analysis

2 weeks - After knowing all the models before and their limitations, we can start modeling our improved version.

Visualization

1 weeks - Once Analysis is complete, come up with visualizations that allow users to quickly understand our findings

Final Paper

2 weeks - Wrap everything up into a final paper

We should keep detailed notes so that we are able to quickly reference back. Always track measurable progress for our progress report. Timeline for each stage adjusted to fit our schedule and deliverables.

Heilmeier Question 8: What are the midterm and final "exams" to check for success? How will progress be measured?