

1           **An exercise in recalibrating perspectives on predictive analytics**

2  
3           FIONA CHIU\*, KIMBERLY GAO\*, and JONATHAN HO\*, Carnegie Mellon University, USA

4  
5           Predictive analytics have become more and more popular and impactful. We note that these have exacerbated inequity and instigated  
6           unfair practices when used unethically and without proper oversight. We endeavour to design a simulation and classroom exercise  
7           that highlights these potential ethical concerns and stimulates critical reflection on identity, decision-making, and the challenges  
8           remaining with ML deployment. In this exercise, we predict academic success and future income of users following the precedents  
9           established in the literature, which also include ways to systematically improve model predictive ability and fairness.

10  
11          **ACM Reference Format:**

12          Fiona Chiu, Kimberly Gao, and Jonathan Ho. 2022. An exercise in recalibrating perspectives on predictive analytics. In . ACM, New  
13           York, NY, USA, 14 pages. <https://doi.org/XXXXXX.XXXXXXX>

14  
15          **1 INTRODUCTION**

16  
17          In recent years, intelligent systems have been designed, developed, and deployed as both augmentations and replacements  
18           of human judgments across countless high-impact decision-making contexts, from social work and criminal justice to  
19           hiring and healthcare. Unfortunately, these AI systems are not necessarily more accurate and equitable than human  
20           decision makers. Much work has discussed how these systems can exacerbate racial biases and strengthen structural  
21           inequalities in already strained yet essential systems, for example Redden et al.'s analysis of child welfare data systems  
22           [10] or Eubanks' book on the impacts of predictive risk models on poor and working-class people in America [5]. In  
23           addition, Veale et al. cover clearly much of the concerns regarding social implications of algorithms [12].

24  
25          Of particular interest is the use of ML to predict student outcomes. Granted, we note that much of these endeavours  
26           are intended to promote student learning and provide proper support to students. Specifically, one such motivation  
27           is that the accurate estimation of students' grades in future courses can inform the design of personalised degree  
28           pathways, as discussed by Polyzou et al. [9].

29  
30          However, much advancement in this field ignores unintended consequences or ethical concerns of predicting student  
31           outcomes. We fail to find research that simultaneously discusses the increasing popularity and subsequent optimisation  
32           of predictive analytics models for student grade prediction. This is especially alarming since these systems are usually  
33           "important, opaque, and destructive" [8], a dangerous trinity. In addition, the Pygmalion effect has been well-documented,  
34           which suggests that teacher expectations are more predictive of student outcomes, but yet can be easily skewed given  
35           an algorithmic prediction of student success.

36  
37          Demonstrably destructive indeed, to use O'Neil's language. In 2020, for example, the U.K. was unable to have British  
38           students sit for advanced-level qualification exams due to the coronavirus pandemic. As reported by Axios, this was  
39           replaced by teachers' estimates of student performance, which were adjusted by an algorithm to ideally compensate for  
40           inflated expected performance [13]. However, students with high grades from less-advantaged schools were more likely

41  
42          \*All authors contributed equally to this research.

43  
44  
45          Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not  
46           made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components  
47           of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to  
48           redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

49          © 2022 Association for Computing Machinery.

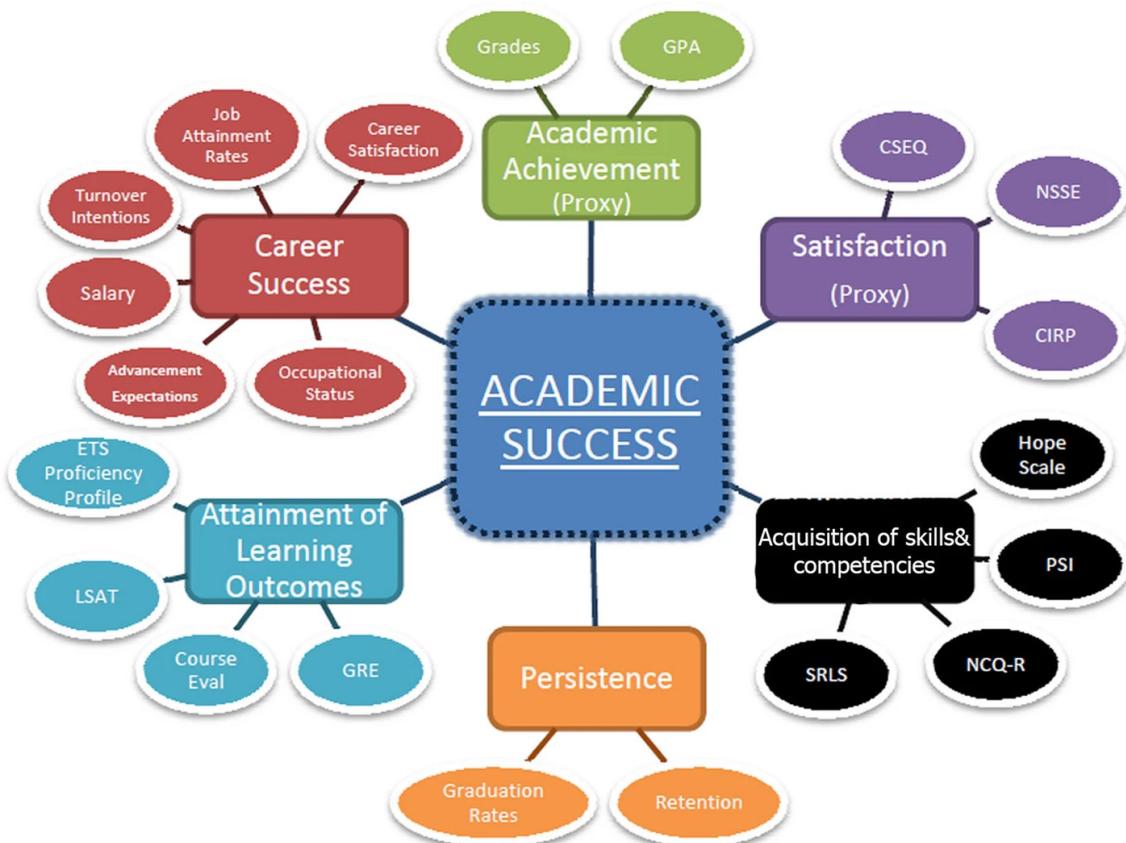
50          Manuscript submitted to ACM

53 to have their scores downgraded, which implies that the algorithm reinforced existing economic and societal biases.  
 54 Similar controversies emerged with the International Baccalaureate Diploma Programme [11], among others.  
 55

56 We move to address this lack of consideration. Through our work, we hope to highlight ethical concerns and rooms  
 57 for abuse with predictive analytics. Subsequently, we spark a guided discussion on protected attributes, historical bias,  
 58 ethical use of ML and data, and social implications of algorithmic decision-support systems.

## 60 2 RELATED WORK

62 Machine Learning Models and Artificial Intelligence have been used extensively in the prediction of an individual's  
 63 academic success, which has several operational definitions that have been explored with a plethora of predictive  
 64 models. York et al. [14] define academic success and its measurements as in Figure 1.



98 Fig. 1. York, Gibson, & Rankin Operationalized Model of Academic Success

101 Similarly, a broad variety of features and their impact on student academic success has been investigated across this  
 102 field. Alyahyan and Düşteğör [1] conducted a literature review that summarized these features in a relational diagram:  
 103

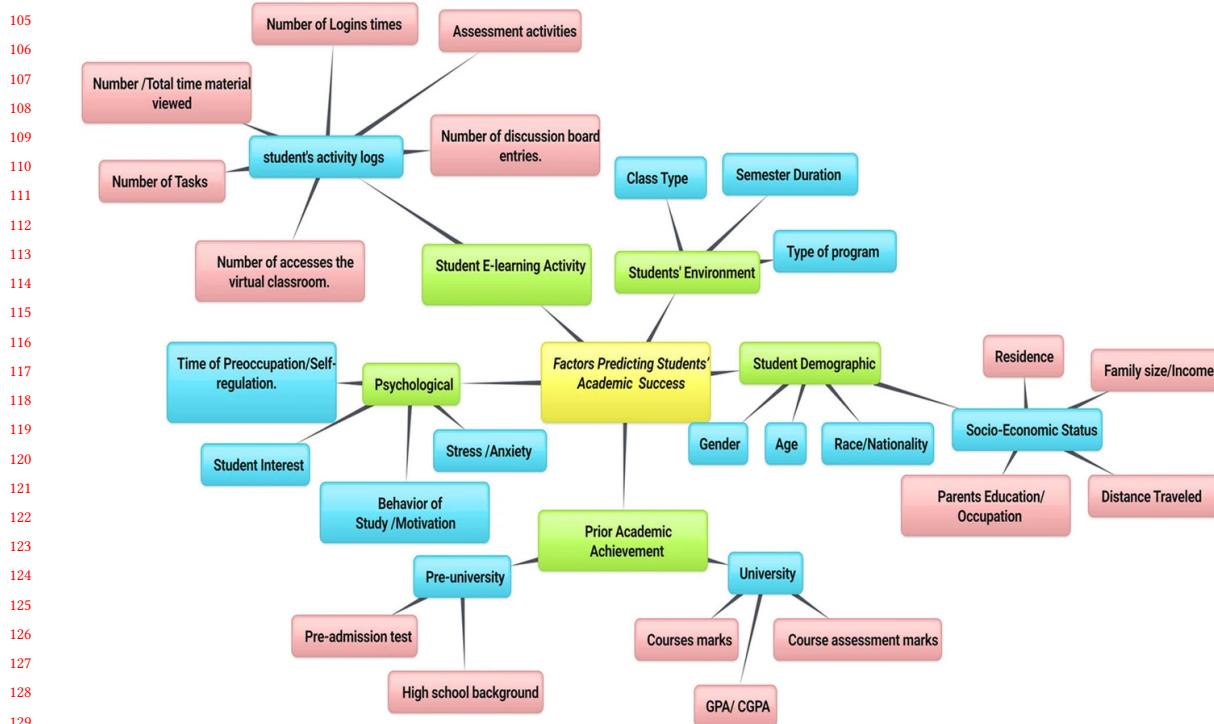


Fig. 2. Relational diagram of features and their impact on student academic success

This project focuses on the outcomes for Persistence (higher education completion) and Career Success (occupational status). The model we aim to build focuses on student demographics and environment which are the most relevant to high school students.

## 2.1 Predicting Student Academic Success in Higher Level Education

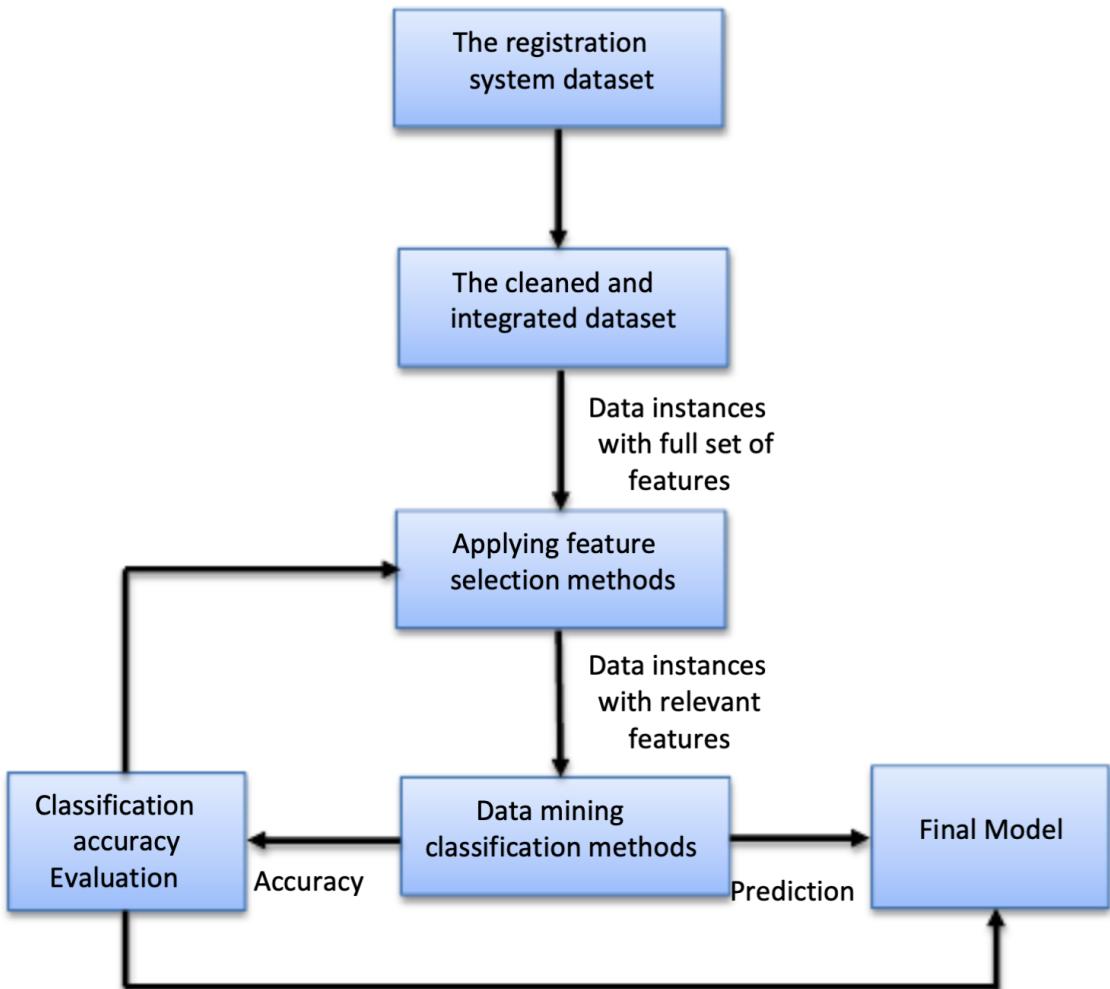
We plan to model the Persistence operationalization of academic success by the binary class classification problem: Does the individual successfully receive a degree in higher education?

Hammoodi et al [6] conducted a study to predict completion in higher education based on socio-demographic information using six prediction approaches: Bagging, DecisionTable, HoeffdingTree, IBK, J48, RandomForest, and RandomTree. The study revealed that societal regression and type of secondary education had the most significant influence on the completion of higher education and that the Bagging Classifier model had the highest accuracy (87.56%). In particular, the exploration and evaluation of prediction approaches for a model based on socio-demographic features is highly relevant to the model we plan to train and provides a starting point for our binary classification model.

The study also used educational data mining (EDM) to determine the socio-demographic features with the most significant impact on completion in HE, which has been greatly attributed to the enhancement of ML model performance according to Ashenafi et al. [2]. EDM is a method or technique to extract significant information that has a potential influence on educational institutes and is often seen as necessary for contemporary education as a large amount of

157 data available on students' databases overrides the ability of a human to extract useful knowledge without applying an  
 158 automated analysis method.

159 Applying the technique used in Hammoodi et al. [6], we employ a similar workflow of feature selection followed by  
 160 applying the prediction model with an accuracy evaluation loop iteration before arriving at the final model.  
 161



199 Fig. 3. Proposed Model for Predicting Student Performance by Hammoodi et al.  
 200

201 The key differentiating factor between Hammoodi et al. [6] and other existing literature is the use of a dynamic  
 202 prediction model rather than a static one which is achieved by running a classification accuracy evaluation loop to  
 203 check if accuracy has dropped (i.e if features have become irrelevant) and replaces them with more relevant features.  
 204 This type of model reflects the dynamicity of present ML research.  
 205

**209 2.2 Equity and Algorithmic Fairness in Student Grade Prediction**

210 In 2021, Jiang and Pardos [7] explored algorithmic fairness and bias in higher education grade prediction with respect  
211 to race to boost equity in education outcome for underserved groups. The researchers identified two major fairness  
212 problems to concentrate on: prediction outcome discrimination from high feature-class correlation as well as prediction  
213 quality disparity from imbalance data. As they trained the model, researchers wanted to maintain accuracy and minimize  
214 overall error which usually leads to optimally fitting the majority group but instead prioritized analysis of high accuracy  
215 across different groups rather than the dataset as a whole. With a dataset from the University of California, Berkeley,  
216 they took student grades from institution wide courses as well as demographics of students at time of admittance as  
217 features to predict future grades. Trying different training methods, researchers found that adversarial learning and  
218 equality based strategies were both effective in improving fairness and predictive performance across race groups.  
219 Equity oriented weighting to help overrepresented groups with historic outcomes unbalanced datasets increased TNRs  
220 and accuracy for these groups. Training the groups separately and excluding sensitive data both decreased accuracy  
221 and performance. Overall, weighting groups and including sensitive attributes will likely require many interactions to  
222 maximize fairness and equity when constructing our model in addition to testing the implementation of adversarial  
223 learning in addition to our predictive classifier.

**224 2.3 Predicting Annual Future Income**

225 In 2006, Dubow et al. [3] established a precedent for using parental occupational status (and other indicators of family of  
226 origin socioeconomic status, or SES) to predict the child's adult occupational outcomes through middle adulthood. Using  
227 longitudinal data from Finland (the Jyväskylä Longitudinal Study of Personality and Social Development, or the JYLS)  
228 and the United States (the Columbia County Longitudinal Study, or the CCLS), the researchers discovered that parental  
229 occupational status is a strong predictor of higher educational status for the child during early adulthood (which also  
230 validates our other branch of exploration), which is in turn a strong predictor of higher occupational status for the  
231 child in middle adulthood, presumably where the child has already settled into a stable career path. Here, occupational  
232 status is defined for 889 specific occupations within 13 categories (for example, executive, sales, or technician). Possible  
233 explanations for this include the tangible opportunities for children's vocational development because of parents with  
234 higher SES. Regardless of the specific mechanics driving such a predictor, however, we seek to apply this relationship to  
235 derive our model predicting future income by way of future occupational success.

**236 3 METHODS: SPECULATIVE FUTURES**

237 Speculative futures provoke, imagine, and dream into what lies ahead, and are inspired by not only current trends but  
238 also art, film, and fiction. We design these futures to spark critical reflections, discussions, and reconsiderations of an  
239 ideal, ethical, and fair society. Through the ensuing debates, we hope to collectively define a preferable future [4].

240 We specifically probe the deployment of predictive analytics systems to predict student life outcomes. We target  
241 New York City residents because as we extend the geographical scope, we lose granularity of data, but yet New York is  
242 diverse enough in and of itself to ensure fairly representative data. We target high school students as the audience of  
243 this exercise for two main reasons: if they were older, there would be nothing to predict — their lives become more and  
244 more silo-ed and deterministic as education levels are realized, jobs are secured, etc. However, if they were younger, they  
245 may not have the necessary worldview and critical faculties to engage in such abstract and philosophical exchanges.  
246 Second, we realize that significant social change best starts from younger generations. Icons like Greta Thunberg and

261 Malala Yousafzai best demonstrate this, and especially with the high demand for new graduates in technology firms,  
 262 the younger generations are well-poised to effect momentous reform and guide innovation.  
 263

264 That being said, we understand that high school students still lack the worldview to critically discuss some popular  
 265 use cases of predictive analytics, such as child welfare or recidivism. To account for this, we focus on contexts that are  
 266 by definition going to be more familiar to users in the education system, namely educational attainment and income.  
 267 This will allow for the maximum relatability and therefore emotional reaction to the proposed scenario.  
 268

#### 269 4 MODEL BUILDING AND EVALUATION

270 We discuss our data collection, predictive model building, model evaluation, and other implementation details more  
 271 in-depth alongside our code , which can be found in the linked repository: <https://github.com/jonathanho168/haii-final-project>.  
 272

273 Here, we provide brief overviews of each model including dataset, features, selected model, and an explanation of  
 274 individual predictions with LIME.  
 275

##### 276 4.1 Model 1: Predicting High School Graduation

277 Using a dataset on Student Performance from 2 Portuguese schools, we select features for mother education level  
 278 (Medu), father education level (Fedu), studytime, and failures based on their correlation and a student's ability to provide  
 279 these values. In training several different models and looking at their F1-scores, we decide on the Linear Support Vector  
 280 Classifier (SVC) as the model for predicting whether or not an individual graduates high school. We ran LIME on two  
 281 individual predictions, one for each of the possible outcomes.  
 282

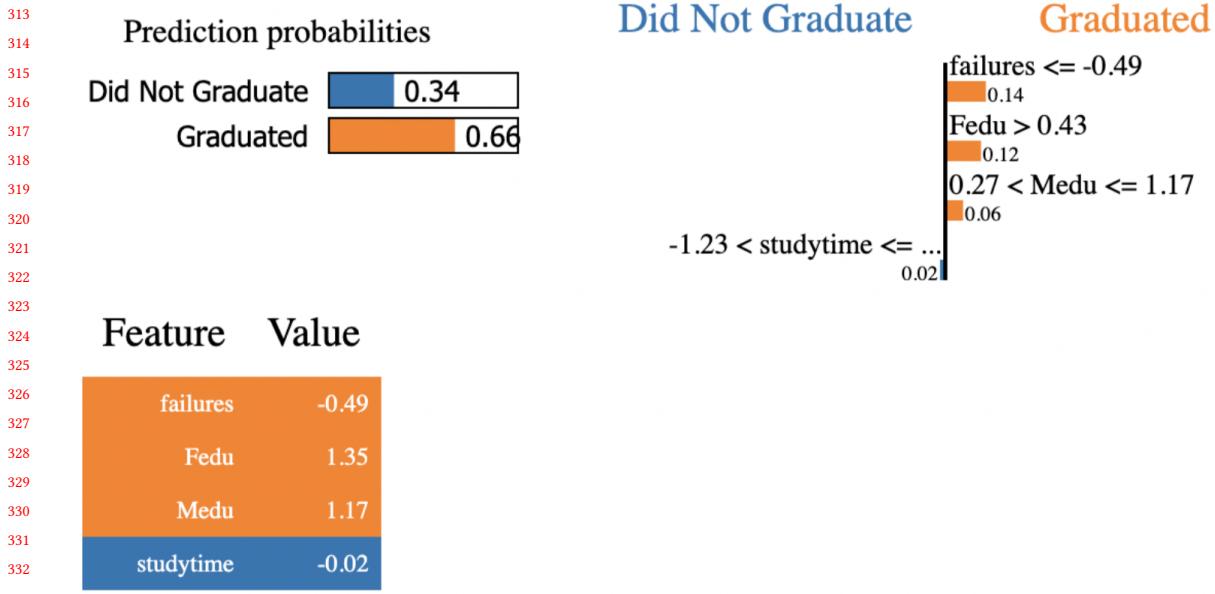
283 The first prediction was for an individual with a Medu of 4, Fedu of 4, study time of 2 and failures of 0. This means  
 284 that both of the individual's parents obtained higher education, their weekly study time was 2 to 5 hours, and they did  
 285 not fail any classes. Based on these inputs, the model predicts that the probability of this individual graduating to be  
 286 0.66. As seen in Figure 4, the model weights the lack of failures with the highest importance as a positive predictor and  
 287 the lower study time as a negative predictor.  
 288

289 The second prediction was for an individual with a Medu of 3, Fedu of 2, study time of 2 and failures of 2. This means  
 290 that the individual's mother completed secondary education, the individual's father completed 5th to 9th grade, their  
 291 weekly study time was 2 to 5 hours, and they failed 2 classes. Based on these inputs, the model predicts the probability  
 292 of this individual not graduating to be 0.70. As seen in Figure 5, the model weights the two failures and the father's  
 293 incompleteness of secondary education to heavily negatively predict the outcome of the student.  
 294

295 From these two predictions, we can see that across the variance of features, the model makes predictions accordingly.  
 296 While study time is the same across the two individuals and they both negatively affect outcome due to the low amount,  
 297 the more important features lie in the disparity of characteristics between the two. When both parents obtain higher  
 298 education, the individual is much more likely to graduate from high school. Lastly, the most important factor is whether  
 299 or not they failed a class.  
 300

##### 301 4.2 Model 2: Prediction SAT Scores

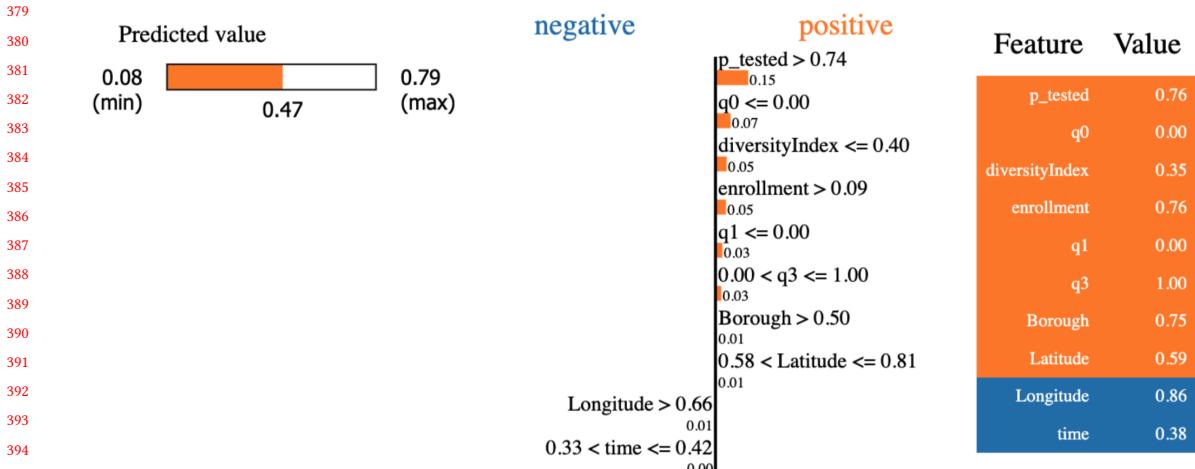
302 The dataset used for this model consisted of information for every accredited high school in New York City (school  
 303 name, borough, building code, street address, latitude/longitude coordinates, phone number, start and end times, student  
 304 enrollment with race breakdown) as well as each schools' average scores on each SAT test section for the 2014-2015  
 305 school year.  
 306



365 Classifier as it had the lowest mean-squared error and highest  $r^2$  value which represents a high level of variability  
 366 being explained by the model.

367 With our scaled data, the predicted value for each datapoint falls on a scale relative to a min and max value. Each  
 368 feature contributes a positive or negative effect on this predicted value. We ran LIME on a sample of individual's schools  
 369 to investigate the feature weights for each prediction.

370  
 371 4.2.1 *Example 1.* The first individual had a predicted value of 0.47. We can see that this is mainly attributed to the high  
 372 percentage of students tested, a diversity index less than 0.40, student enrollment higher than 0.09 and having words in  
 373 the school name that fall in q3 (the vocab quartile with the highest average SAT score). The main negative attribute  
 374 includes having time spent in school less greater than 0.33.  
 375  
 376  
 377  
 378



396 Fig. 6. LIME Explanation for Prediction of Model 2, Example 1  
 397  
 398  
 399  
 400

401 4.2.2 *Example 2.* We can see that the second individual has a predicted value of 0.15. The strongest weighted negative  
 402 attributes include having a diversityIndex greater than 0.57, percent tested less than 0.74 and having no words int their  
 403 school name in the highest quartile. This is further amplified by time spent in school being greater than 0.33 and having  
 404 words in their school name in the two lowest quartiles (q0 and q1). Compared to individual 1, although this individual's  
 405 school has a significantly lower score in student enrollment, it is still not weighted as heavily as percent tested and  
 406 diversityIndex which have threshold values that seem to be more normally distributed, therefore, student enrollment is  
 407 actually a positively weighted feature in this case.  
 408  
 409

410 4.2.3 *Example 3.* The third individual investigated had a predicted value of 0.73. This was mainly attributed to the fact  
 411 that they had a very high percent tested (0.97) and low diversityIndex (0.21). In general, their other attributes, shorter  
 412 time spent in school, having words in their school name in the highest quartile but not the lowest 2 quartiles, and  
 413 enrollment greater than 0.10, all contribute to the very high positive score.  
 414  
 415

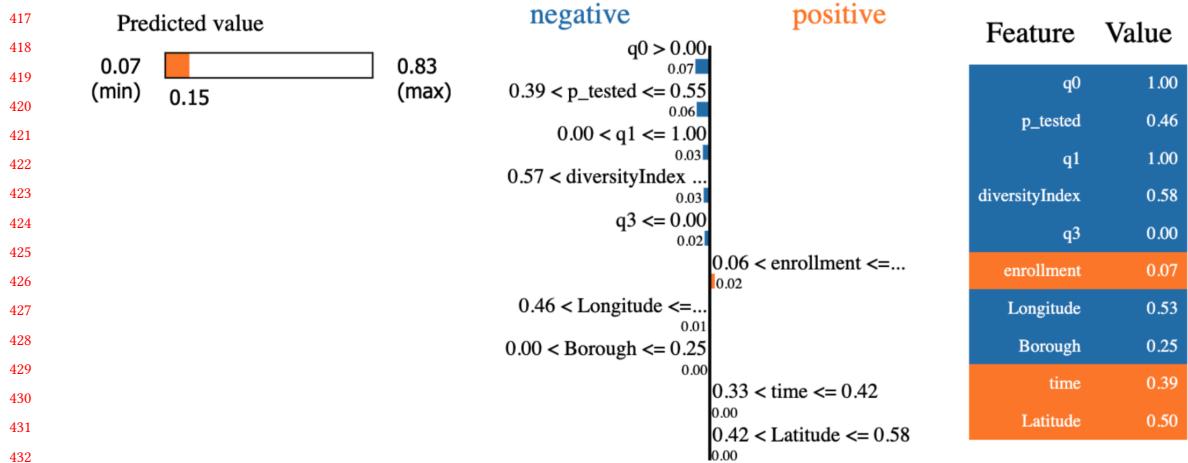


Fig. 7. LIME Explanation for Prediction of Model 2, Example 2

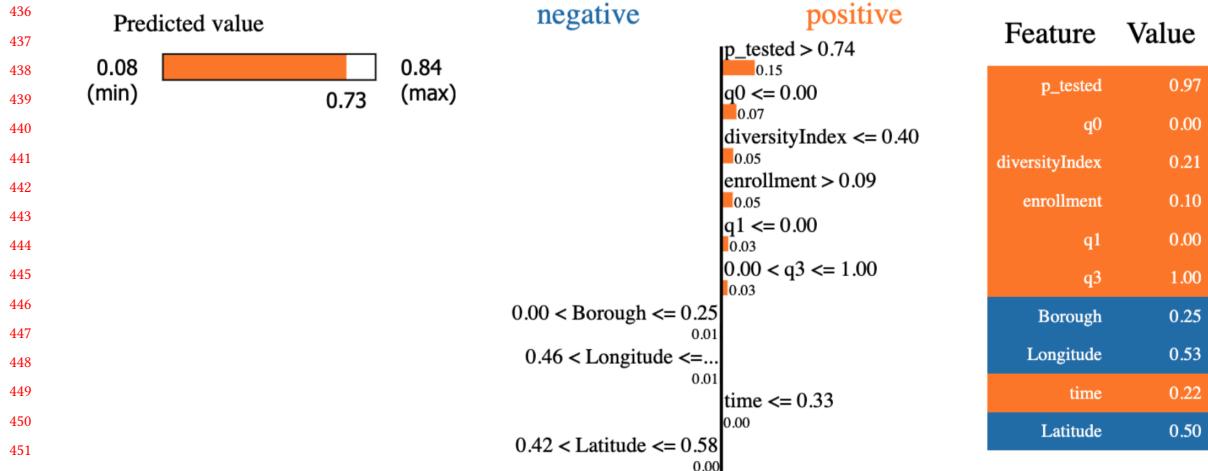


Fig. 8. LIME Explanation for Prediction of Model 2, Example 3

#### 4.3 Model 3: Predicting Student Enrollment in College

The dataset used for this model consisted of sociodemographic data (gender, residence, parent\_age, parent\_salary, house\_area, average\_grades, whether or not the parent was in college) and school data (accreditation, type), as well as whether the student enrolled in college.

To construct an accurate model to predict whether a student will be admitted and enroll in college, we trained and evaluated several models using the following machine learning techniques:

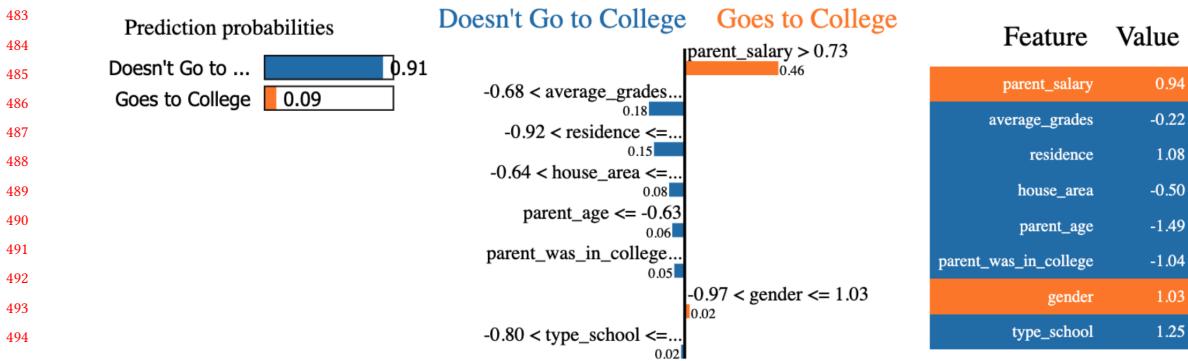
- Logistic Regression
- K-nearest Neighbors
- Random Forests

469     • SVM

470     The SVM and Random Forest Classifier had the highest accuracy and f1\_score but the SVM had a slightly lower  
 471     False Positive score so we decided to select this for our final model. After initial investigation into feature importance,  
 472     we found that the selected model weighted parent\_salary, house\_area, and average\_grades to be the best predictors of  
 473     whether or not a student will enroll in college. To determine the interpretability of our model we ran LIME on a sample  
 474     of individuals to investigate the feature weights for each prediction.

475     4.3.1 *Example 1.* The first individual was predicted not to go to college with a probability of 91%. This was largely  
 476     attributed to their average\_grade being less than 0.65, their residence score being greater than -0.92, and their house  
 477     area being less than 0.09. Although their parent\_salary being greater than 0.73 seemed to contribute significantly to a  
 478     positive prediction, this wasn't enough to outweigh the strength of the negative predictors.

479

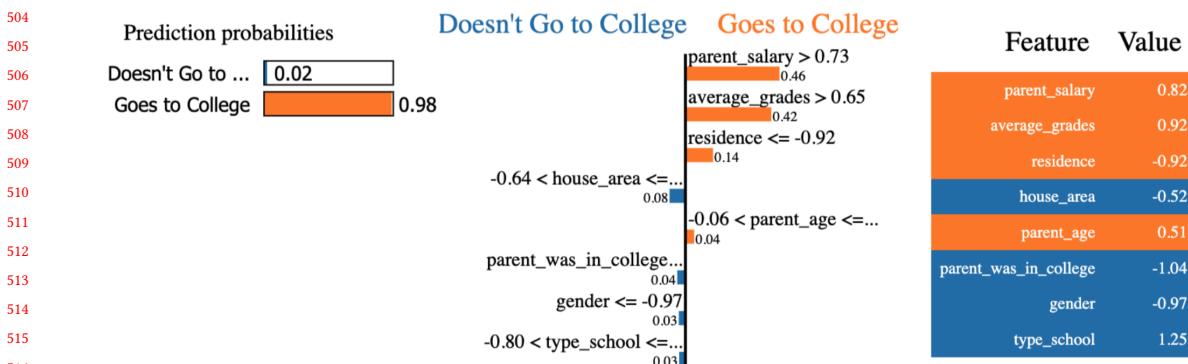


496     Fig. 9. LIME Explanation for Prediction of Model 3, Example 1

497

498     4.3.2 *Example 2.* The second individual investigated was predicted to go to college. We can see that this is highly  
 499     attributed to their parent salary being higher than 0.73 and their average\_grade being higher than 0.65. Although their  
 500     scaled house area contributed to a negative prediction this was not enough to outweigh the positive predictors.

501



517     Fig. 10. LIME Explanation for Prediction of Model 3, Example 2

518

519

520

521    4.3.3 *Example 3.* From example 1 and 2 we were able to see the effect of the three strongest predictors; parent\_salary,  
 522    average\_grades, and house\_area. In order to differentiate the strength of these predictors we pick out the following  
 523    case. This individual is predicted to go to college with a 97% probability. Although their average grade of -0.87 is quite  
 524    low and less than the threshold of -0.68 and contributes significantly to a negative prediction, we see that the high  
 525    parent\_salary, house\_area, and residence score almost completely outweighs the negative predictor.  
 526

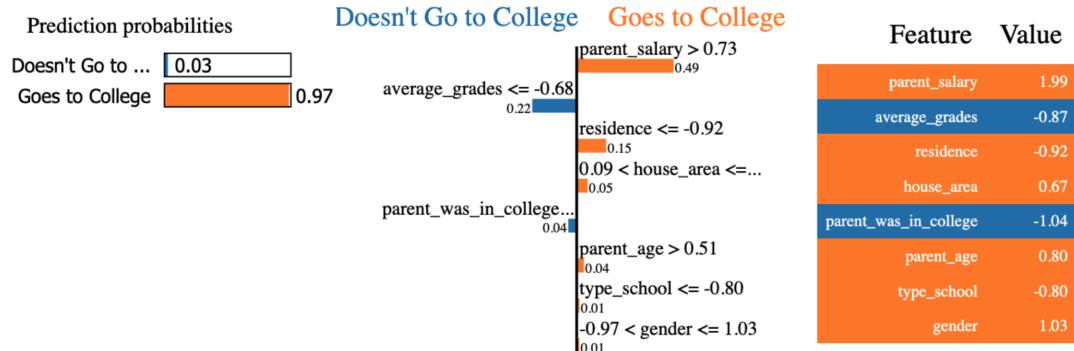


Fig. 11. LIME Explanation for Prediction of Model 3, Example 3

## 5 APPLICATIONS IN AN EDUCATIONAL CONTEXT

546    We anticipate and expect our system as a whole to help students better understand political, social, and technical  
 547    realities that ethics interventions must navigate. Hopefully, with these considerations in mind, students will be better  
 548    equipped to connect ideation and implementation of novel systems with downstream use, perhaps even concretely  
 549    operationalising and ensuring ethical use.  
 550

551    That being said, the specific use cases and deployments of our system as an educational tool are dependent on the  
 552    learning environment, the abilities of the teacher, and the level of the student users. In addition, we are not experts in  
 553    learning sciences, nor do we pretend to have nontrivial pedagogical experience. Therefore, we leave the design of the  
 554    education exercise to the reader.  
 555

556    However, we do share our expectations of the system as it is used to highlight ethical issues.

557    Prior to use of the system, we would like students to engage in a discussion of identity – how are students defining  
 558    themselves, and what aspects of their background do they highlight most? We would like the discussion to specifically  
 559    focus on protected attributes, such as race and gender, as they play into our lifestyles, personality, etc. Furthermore, we  
 560    would also like to have discussions on whether students would like to be judged based on those attributes – why or  
 561    why not? This series of questions is prime breeding ground for cognitive dissonance, and some reasonable discomfort  
 562    is expected.  
 563

564    As discussed in our implementation, we also make specific choices – for example, we do not balance the data, nor  
 565    do we normalize our continuous variables. It would be a good exercise for the instructor to pose different questions  
 566    regarding data collection and use – for example, with respect to data balancing, one could pose the following: “sometimes  
 567    the world just has unbalanced data. There are more microwaves than toasters out there, so would we want to have  
 568    a dataset that has just as many toasters and microwaves?” Obviously, some exposure to data science practices and  
 569    vocabulary would be helpful.  
 570

573 After the exercise, we would like students to revisit the same questions they considered earlier. In addition, we would  
574 like students to articulate clearly the reasons behind their discomfort with such a predictive model and its proposed use.  
575 Expecting a reasonably diverse set of answers, perspectives, and experiences, we hope to also scaffold a discussion on  
576 what attributes are ethically usable and how to enforce ethical design and deployment.  
577

578 It seems reasonable to the authors to position this exercise in advance of a research-intensive project as well. Students  
579 could use this exercise to surface core beliefs and explore advances and industry practices that they would like to see  
580 implemented, revised, or even eradicated. We hope that students will also be empowered and excited to take action,  
581 perhaps even pursuing more human-centered fields later on.  
582

## 583 584 585 6 USER STUDIES

586 In this exercise, we first discussed identity, as mentioned above. Then, we transitioned to model interactions. We first  
587 allowed users to both review the raw datasets we use peruse some preliminary exploratory data analysis (provided)  
588 and the attributes included, such as demographic features and student grades. We allowed them to examine also what  
589 features we ended up using, and allowed them to ruminate over the justifications (if they exist) behind this selection.  
590

591 We then made predictions based on user survey responses (regarding their background, etc.), and asked the user to  
592 imagine being judged in real life based on these predictions.  
593

594 After this examination of the simulated predictions, we returned to the original questions.

595 Through this process, our first participant was concerned about the inclusion of ethnicity as a consideration. However,  
596 upon probing, a discrepancy emerged: the participant used ethnicity to form a personal sense of identity, but did not  
597 want others to do the same. We found that the issue at hand is not specifically the use of ethnicity, but the use of a  
598 potentially ambiguous and subjective feature. Here, this participant believed that ethnicity is subject to stereotypes,  
599 skewed expectations and biased preconceptions by others.  
600

601 Rather, the participant proposed basing a profile on more obvious and salient features, such as ones easily verifiable  
602 – the examples provided were the size of one's eyes or their gender, which are arguably less nuanced than ethnicity.  
603 Of course, this was in the case that a human were judging (hence, the participant also mentioned compatibility in  
604 interpersonal relationships as judged by others as a factor of identity), so this concern does not necessarily extend to  
605 our system. That being said, we still remove ethnicity as a predictor and as a feature from our models (specifically  
606 model 3), to focus the discussion and remove distractors.  
607

608 Another user included hometown, age, languages spoken, and reading habits when initially asked about forming  
609 identity. This seemed random to the interviewers, so further probing established that the user considered these attributes  
610 to define and facilitate the user's daily interactions and activities. Using the system and seeing the predictions it made  
611 raised discomfort, but the user was unable to clearly identify the issue. An attempt was made, however, and the user's  
612 response suggested concerns with the unrepresentative and context-agnostic predictions that did not take into account  
613 many factors of growth and education, such as social skills. Probing on what features would be valid for such a system  
614 was not productive, however – there were too many points of controversy.  
615

616 From these user studies, it seems that our system effectively raised concerns about the formation and use of identity  
617 profiles, and these users were critically reflecting on how they as peers and citizens make judgements. It follows  
618 reasonably that with input from learning experts, an effective exercise can be refined and deployed in learning contexts.  
619

620 In addition, from all of the users we interviewed, all raised concerns on the possible misuse of the system and the  
621 validity of the predictions presented. This leads us to design the following ethical restrictions on system use.  
622

## 625 7 ETHICAL RESTRICTIONS ON SYSTEM USE

626 It is apparent to the authors that this system could very easily be taken to make actual predictions. It would be ironic  
627 and hypocritical if we did not encode any mechanisms to prevent said abuse of our system. Hence, we propose the  
628 following:

629 First, all users must be registered with an email address tied to an educational or academic institution. Second, all  
630 users, once registered and verified to be in an instructional position, must sign a contract to prevent inappropriate use  
631 of this system. All use cases not explicitly permitted must be granted explicit permission in writing from a majority of  
632 the creators.

633 Third (and perhaps most important), is that we do not store any user data. All predictions are derived from existing,  
634 publicly available datasets and statistics. We do not know who will be using our system, nor do we have any methods  
635 for gleaning that information. The risk of security breaches and loss of personally identifiable information is therefore  
636 negligible.

637 The authors disclaim that they will need to consult with legal experts to determine an ideal realization of said  
638 restrictions and use conditions.

## 643 8 CONCLUSION

644 In response to the overwhelming increase in the use of predictive analytics systems and the subsequent increase in  
645 ethical concerns regarding their deployment, we design an educational exercise to highlight these issues and spur  
646 ethical considerations. By building predictive models based on publicly available datasets, we are able to simulate living  
647 under the effect of these predictions, and these stimulate strong responses and allow for a critical discussion on the  
648 validity and fairness of these systems, and how these systems can be improved.

## 653 REFERENCES

- 654 [1] Eyman Alyahyan and Dilek Düşteğör. 2020. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education* 17, 1 (10 Feb 2020), 3. <https://doi.org/10.1186/s41239-020-0177-7>
- 655 [2] Michael Mogessie Ashenafi, Giuseppe Riccardi, and Marco Ronchetti. 2015. Predicting students' final exam scores from their course activities. In *2015 IEEE Frontiers in Education Conference (FIE)*, 1–9. <https://doi.org/10.1109/FIE.2015.7344081>
- 656 [3] Eric F. Dubow, L. Rowell Huesmann, Paul Boxer, Lea Pulkkinen, and Katja Kokko. 2006. Middle childhood and adolescent contextual and personal predictors of adult educational and occupational outcomes: A mediational model in two countries. *Developmental Psychology* 42 (2006), 937–949. <https://doi.org/10.1037/0012-1649.42.5.937>
- 657 [4] Anthony Dunne and Fiona Raby. 2014. *Speculative everything: Design, fiction, and Social Dreaming*. MIT.
- 658 [5] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, Inc., USA.
- 659 [6] Mahmood Hammoodi and Ahmed Al-Azawei. 2022. Using Socio-Demographic Information in Predicting Students' Degree Completion based on a Dynamic Model. *International Journal of Intelligent Engineering and Systems* 15 (02 2022), 2022. <https://doi.org/10.22266/ijies2022.0430.11>
- 660 [7] Weijie Jiang and Zachary A. Pardos. 2021. Towards Equity and Algorithmic Fairness in Student Grade Prediction. *CoRR* abs/2105.06604 (2021). arXiv:2105.06604 <https://arxiv.org/abs/2105.06604>
- 661 [8] Cathy O'neil. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- 662 [9] Agoritsa Polyzou and George Karypis. 2019. Grade prediction with course and student specific models. *CoRR* abs/1906.00792 (2019). arXiv:1906.00792 <http://arxiv.org/abs/1906.00792>
- 663 [10] Joanna Redden, Lina Dencik, and Harry Warne. 2020. Datafied child welfare services: unpacking politics, economics and power. *Policy Studies* 41, 5 (2020), 507–526. <https://doi.org/10.1080/01442872.2020.1724928> arXiv:<https://doi.org/10.1080/01442872.2020.1724928>
- 664 [11] Tom Simonite. 2020. Meet the secret algorithm that's keeping students out of college. <https://www.wired.com/story/algorithm-set-students-grades-altered-futures/>
- 665 [12] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3173574.3174014>

- 677 [13] Bryan Walsh. 2020. How an AI grading system ignited a national controversy in the U.K. [https://wwwaxios.com/2020/08/19/england-exams-](https://wwwaxios.com/2020/08/19/england-exams-algorithm-grading)  
678 [algorithm-grading](#)
- 679 [14] Travis T. York, Charles Gibson, and Susan Rankin. 2015. Defining and measuring academic success. *ScholarWorks@UMass Amherst* (2015).  
680 <https://doi.org/10.7275/hz5x-tx03>

681 Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009  
682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728