

Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data

by Renzhe Yu, Qiujie Li, Christian Fischer, Shayan Doroudi and Di Xu

presented @ EDM 20

Roadmap

1

Why
should we care?

2

What
are the contributions?

3

How
do they do it?



1. Motivation

MOTIVATION

Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner



MOTIVATION

Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner



MOTIVATION

Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner

Unfortunately, our **society is historically biased against minorities**, and we do not want to make use of a model that could eventually amplify existing achievement gaps.



MOTIVATION

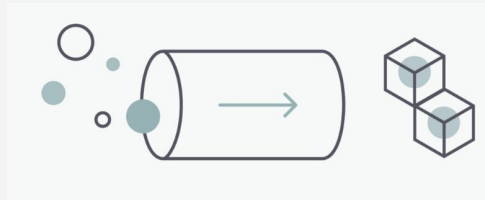
Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner



Unfortunately, our **society is historically biased against minorities**, and we do not want to make use of a model that could eventually amplify existing achievement gaps.

A academic institution will look to make decisions that are both **useful** and **fair**.



MOTIVATION

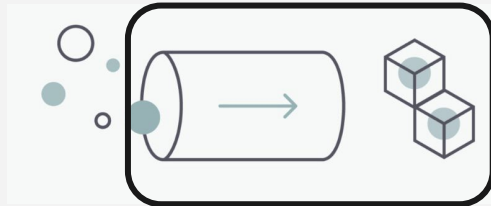
Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner



Unfortunately, our **society is historically biased against minorities**, and we do not want to make use of a model that could eventually amplify existing achievement gaps.

A academic institution will look to make decisions that are both **useful** and **fair**.



MOTIVATION

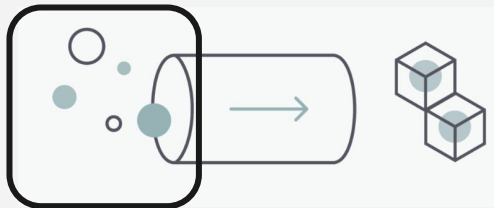
Education benefits from AI-based approaches for example to

- predict graduation grade
- identify students who need support in a timely manner



Unfortunately, our **society is historically biased against minorities**, and we do not want to make use of a model that could eventually amplify existing achievement gaps.

A academic institution will look to make decisions that are both **useful** and **fair**.





2. Contributions

CONTRIBUTIONS

Using real-life data collected from a large public american university, the authors

- show how to evaluate the **costs and benefits of utilizing different realistic data sources** across both short- and long-term prediction tasks
- **recommend** to include data about the **learning behavior** of the students
- reflect on the propagation of data bias on the prediction results.

The most valuable contribution of the study lies within the **knowledge systemization effort**. The authors give a lot of insight on the trade-offs offered to academic institutions when using common classification algorithms.



3. Going in depth

DATA SOURCES

The dataset is composed of information from **2,093 students** from ten online, introductory STEM courses taught from 2016 to 2018 at a university in the US. Six of the courses were in **public health** while the remaining four were distributed across **biology, chemistry and physics**.

Institutional	Click	Survey
Female	Total clicks	Effort regulation
Transfer	Total clicks by category	Time management
Low income	Total time	Environment management
First-gen	Total time by category	Self-efficacy
URM	(All above for the first 5 weeks)	
SAT total score		
High school GPA		

DATA SOURCES

The dataset is composed of information from **2,093 students** from ten online, introductory STEM courses taught from 2016 to 2018 at a university in the US. Six of the courses were in **public health** while the remaining four were distributed across **biology, chemistry and physics**.

Institutional	Click	Survey
Female 72%	Total clicks	Effort regulation
Transfer	Total clicks by category	Time management
Low income 48%	Total time	Environment management
First-gen 54%	Total time by category	Self-efficacy
URM 33%	(All above for the first 5 weeks)	
SAT total score		
High school GPA		

METHODS

Tasks:

- Short-term: is the student final grade above the class median?
- Long-term: is the student GPA above the class median?
- Model should span across courses: 9 courses used to train, the remaining course used as test.

Model: Best out of logistic regression, SVM and random forests.

Metrics:

- Usefulness = Accuracy
- Fairness = { Disparity in accuracy, FPR (“overestimate students”) and FNR (“underestimate students”) }

RESULTS

Feature	Accuracy		FPR		FNR	
	Short	Long	Short	Long	Short	Long
Institutional	0.618	0.599	0.467	0.412	0.299	0.389
Click	0.602	0.613	0.485	0.385	0.313	0.389
Survey	0.534	0.557	0.599	0.385	0.336	0.502
Institutional+Click	0.670	0.650	0.351	0.330	0.310	0.370
Institutional+Survey	0.633	0.608	0.398	0.397	0.337	0.386
Click+Survey	0.609	0.604	0.431	0.457	0.353	0.335
Institutional+Click+Survey	0.675	0.638	0.348	0.402	0.303	0.323

RESULTS

	Feature	Accuracy		FPR		FNR	
		Short	Long	Short	Long	Short	Long
VS	Institutional	0.618	0.599	0.467	0.412	0.299	0.389
	Click	0.602	0.613	0.485	0.385	0.313	0.389
	Survey	0.534	0.557	0.599	0.385	0.336	0.502
	Institutional+Click	0.670	0.650	0.351	0.330	0.310	0.370
	Institutional+Survey	0.633	0.608	0.398	0.397	0.337	0.386
	Click+Survey	0.609	0.604	0.431	0.457	0.353	0.335
	Institutional+Click+Survey	0.675	0.638	0.348	0.402	0.303	0.323

Combining sources is usually **helpful**. The most noticeable accuracy increase is obtained with the combination of institutional + click.

RESULTS

Feature	Accuracy		FPR		FNR	
	Short	Long	Short	Long	Short	Long
Institutional	0.618	0.599	0.467	0.412	0.299	0.389
Click	0.602	0.613	0.485	0.385	0.313	0.389
Survey	0.534	0.557	0.599	0.385	0.336	0.502
Institutional+Click	0.670	0.650	0.351	0.330	0.310	0.370
Institutional+Survey	0.633	0.608	0.398	0.397	0.337	0.386
Click+Survey	0.609	0.604	0.431	0.457	0.353	0.335
Institutional+Click+Survey	0.675	0.638	0.348	0.402	0.303	0.323

Combining sources is usually **helpful**. The most noticeable accuracy increase is obtained with the combination of institutional + click.

Click data is especially useful for the long-term tasks.

RESULTS

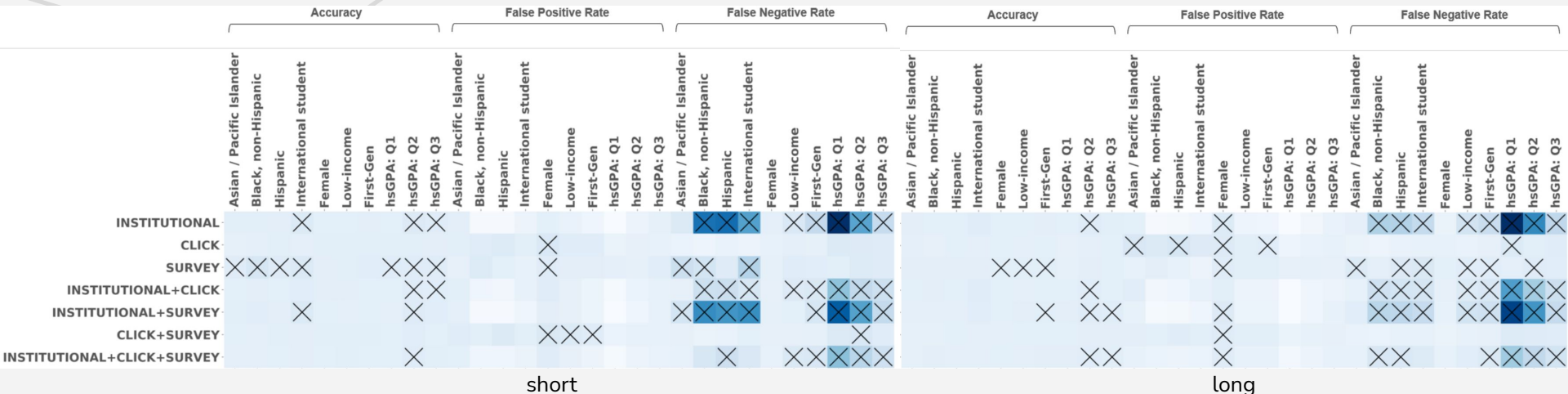
Feature	Accuracy		FPR		FNR	
	Short	Long	Short	Long	Short	Long
Institutional	0.618	0.599	0.467	0.412	0.299	0.389
Click	0.602	0.613	0.485	0.385	0.313	0.389
Survey	0.534	0.557	0.599	0.385	0.336	0.502
Institutional+Click	0.670	0.650	0.351	0.330	0.310	0.370
Institutional+Survey	0.633	0.608	0.398	0.397	0.337	0.386
Click+Survey	0.609	0.604	0.431	0.457	0.353	0.335
Institutional+Click+Survey	0.675	0.638	0.348	0.402	0.303	0.323

Combining sources is usually **helpful**. The most noticeable accuracy increase is obtained with the combination of institutional + click.

Click data is especially useful for the long-term tasks.

Survey data is not very useful, it often overestimates (nice try students) and only marginally helps with some of the prediction errors.

RESULTS



Combining sources somewhat helps but **does not guarantee fairness**.

Institutional data causes prediction to be **based on the majority class** within the subpopulation (underestimates underrepresented students and overestimates women) -> class imbalance affects fairness.

Identity-blind click data is not free from bias (FPR), and does not “debiase” institutional data.

Survey data is not fair.

SIDE NOTE

What if we simply **remove a specific institutional feature** (e.g., gender), would it eliminate the bias against the corresponding disadvantaged group (e.g., female)?

No, no matter the feature.

This could be explained by the typical **intersectionality** of minority identities, i.e., a first-generation college student is most likely to come from a low income family.

CONCLUSION

We have seen a methodology to evaluate data source and pick the one(s) offering the best performance.

- ★ Combining several data sources helps towards a better accuracy of results.
- ★ Using different data sources does not guarantee fairness.
- ★ Adding LMS (behavioral data such as clicks) improves the usefulness and fairness of predictions for this dataset.
- ★ Self-assessed samples are not yet valuable, more research is needed to build better surveys.