Explainable Machine Learning for Public Policy

UC Berkeley School of Information
MIMS Final Project 2018

Samuel Meyer   Shrestha Mohanty
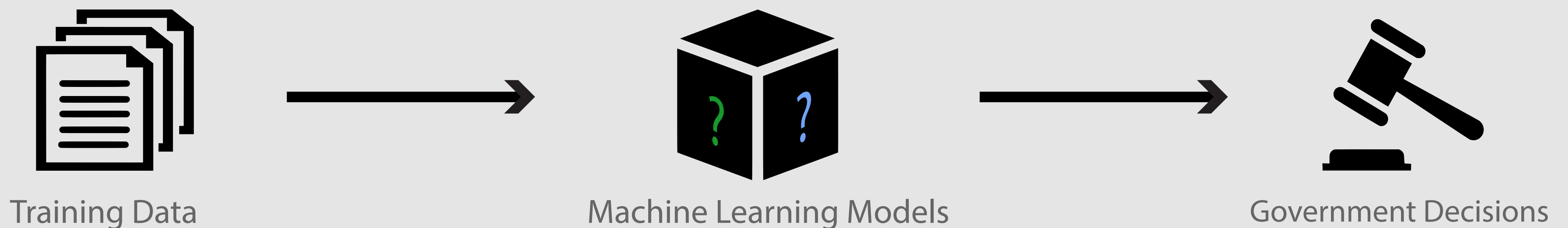Sung Joo Son   Monicah Wambugu
Advisor: David Bamman

Don't let the machine (learning) tell you what to do.

Instead, use the Machine Learning Explorer to tell the algorithms what to do. Governments are looking to use machine learning to make decisions about citizens. But citizens have a right to know how those decisions are made, and to argue to change those processes. The Machine Learning Explorer makes that possible for decisions implemented by machine learning.
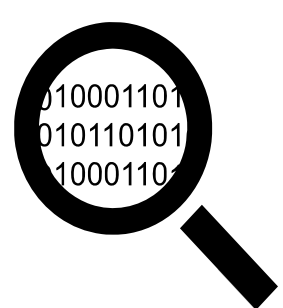
Normally, public policy is publicly debated, and any agencies that implement policies post details of how they make those decisions. Similarly, when a judge makes a decision, she writes an opinion explaining how she made the decision so others can review it later.

Events in the World → Public Debates → Government Decisions

When machine learning is used to make decisions, the training data is not released because it often contains sensitive information about citizens, and the details of the decision-making process are hidden.

Training Data → Machine Learning Models → Government Decisions

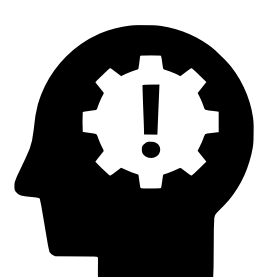## But with the Machine Learning Explorer, you can

Inspect
how the machine learning works

Communicate
with other citizens and the authorities

Improve
the algorithmic decisions

Accuracy: 75.76%

Confusion Matrix:
Predictions for all

	Predicted negative	Predicted positive
Actually negative	True negative 35 (35.4%)	False positive 14 (14.1%)
Actually positive	False negative 10 (10.1%)	True positive 40 (40.4%)

True Positive Rate: 74.1%

Factor	If factor is true, *Final Grade: Passing* is — Less likely ←	→ More likely	Odds Ratio ?			
▼ More classes failed			0.46			
◑ ▼ Home to school: more than 1 hour			0.54			
▼ Higher Grade At End of Last Month			1.33	✏	Balance Model	0.2864
◑ ▼ Home to school: 30 min - 1 hour			0.75	✏	Balance Model	-0.2815
◑ ▼ Home to school: 15 - 30 min			0.78	2	Balance Model	-0.2525
◑ ▼ Sex: Male (not Female)			0.92	✏	Balance Model	-0.0782
▼ More absenses			0.99	3	Balance Model	-0.0102
			0.28		Balance Model	-1.2555

"Final Grade: Passing"
Correctly predicted - 40 / 54
Incorrectly predicted - 14 / 54

Not "Final Grade: Passing"
Correctly predicted - 35 / 45
Incorrectly predicted - 10 / 45