



SUSTAINABLE DEVELOPMENT GOALS



The SDGs, why should I care?

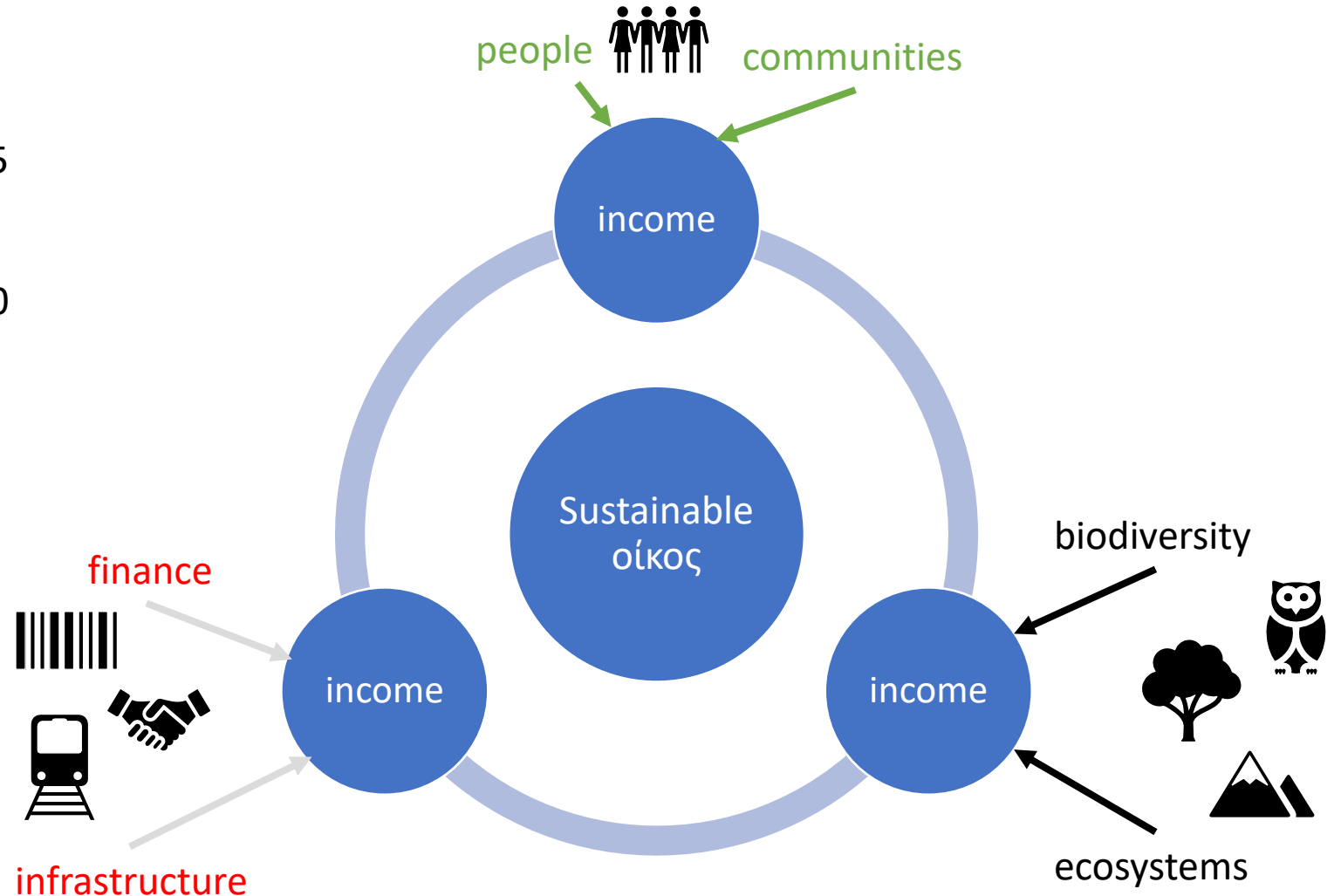
Am I already contributing?

David Lusseau

@lusseau

Current consensus goals: find sustainability for our planetary socioecological systems

Rio declaration 1992
↓
Millennium Development Goals 2000-2015
↓
Sustainable Development Goals 2015-2030



Aligning with the SDGs – how are we helping?

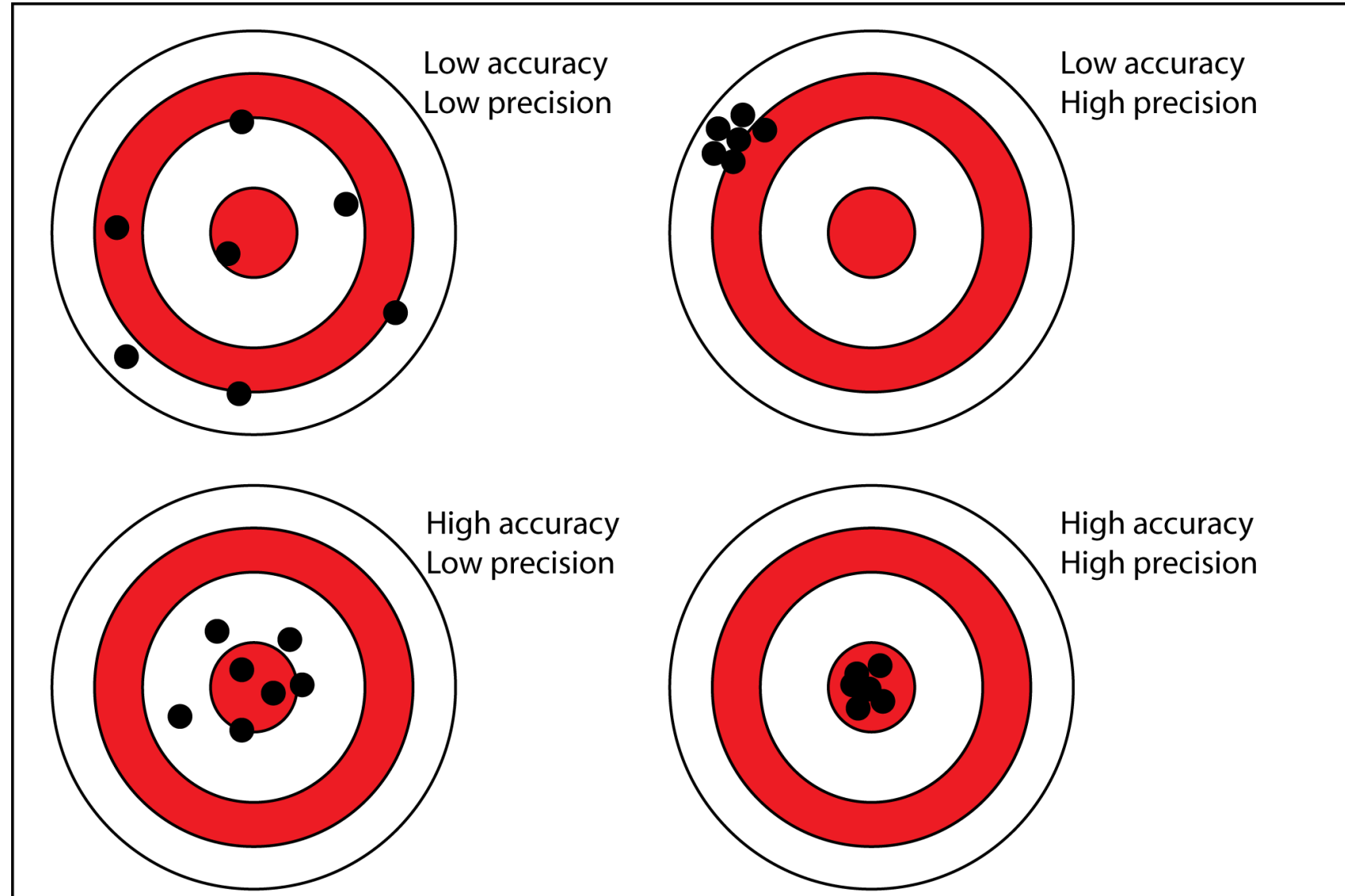
- Landscape activities on the SDGs
- Can we categorise text to the SDG labels?
- Machine learning approach (neural network multi-label classification of text) – ‘shallow’ deep learning

All models are wrong,
but some are useful

Unsupervised learning models are tools

- These are useful to categorise large ensemble of text
 - Landscape the activities of a company
 - Landscape the contributions of universities/departments
 - Landscape learning outcomes of courses
- For precise/high resolution estimates, consult your friendly SDG researcher
 - i.e., “is this particular article contributing more to SDG1 or SDG10?”
 - i.e., “to which SDG target does this research objective contribute?”

End of disclaimer

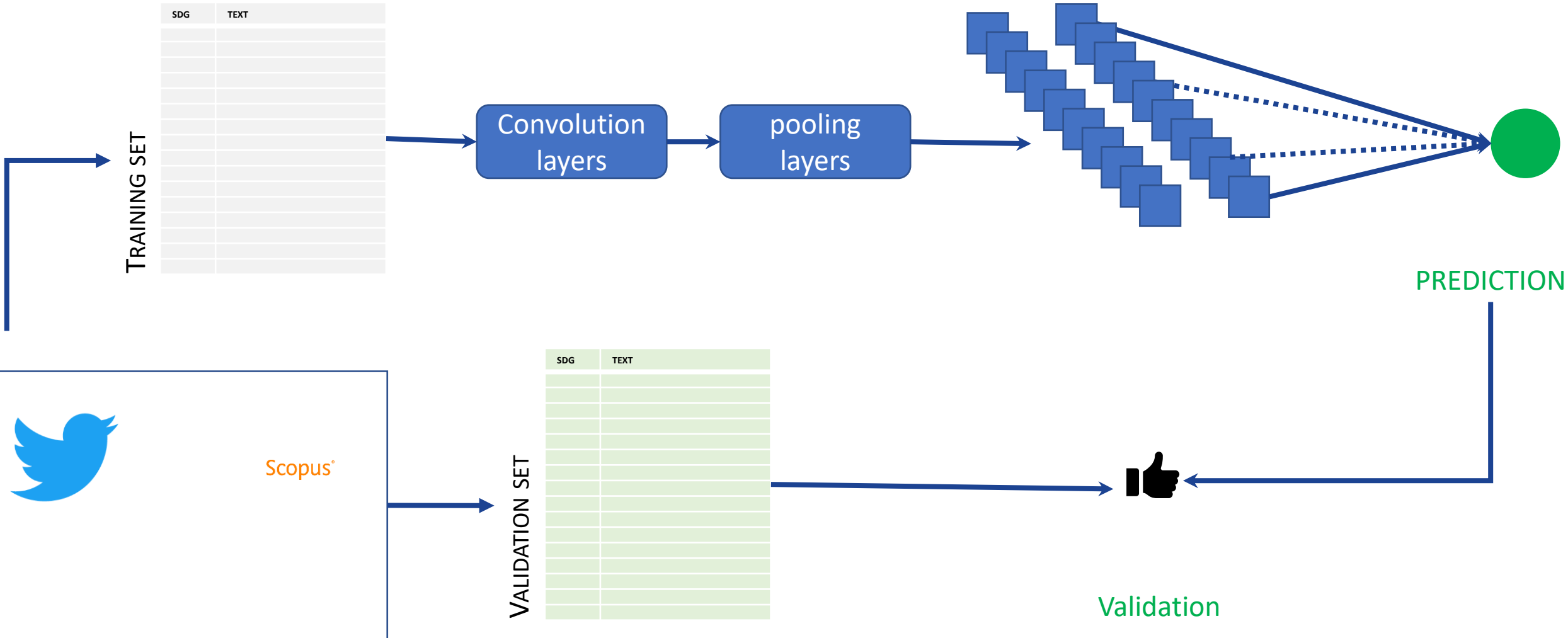


Training a deep-learning model



- Pipeline off Twitter
- All tweets containing “sdg1” to “sdg17”
- Censoring: keep only tweet mentioning *one and only one* sdg
 - At moment $\sim \frac{1}{4}$ million
- Text cleaning, emoticon/emoji translation, deal with special characters, stemming

Convolutional neural network - fitting



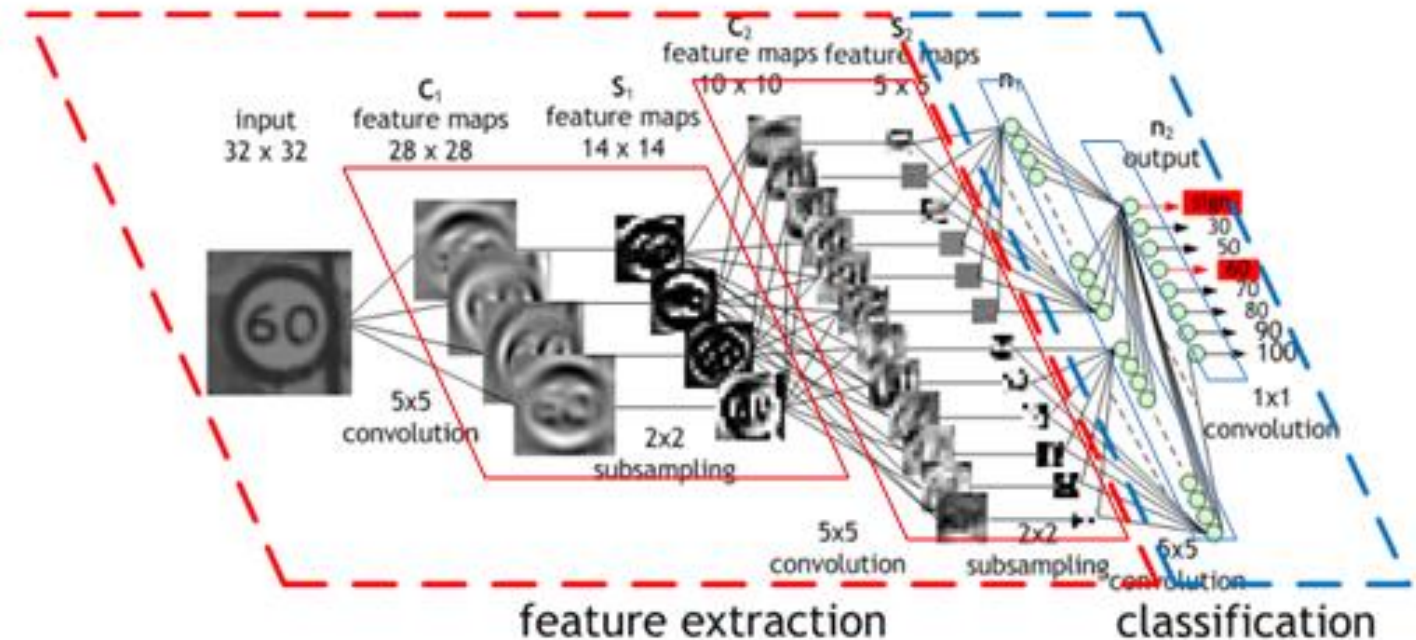
Accuracy ~96% on training set, ~93% on validation set

'shallow' deep-learning

- Current model ensemble:
 - Trained on 80% of text and validated on 20% of text
 - (blocked random sampling to ensure coverage of all sdgs)
 - 1 CN, 1 max pooling layer, 3 full layers (including the last one outputting to SDG labels) – tried up to >20 layers, simpler performs better.
 - Fitting on multiple models (with variation on hyperparameters and replication across validation set subsetting within models)
- Predictions on new text
 - Retain predictions with confidence >90% (arbitrary + remember in ~7% of cases this precise prediction will be inaccurate)
 - Mode of retained label SDG prediction is the SDG category for the output
 - conservative

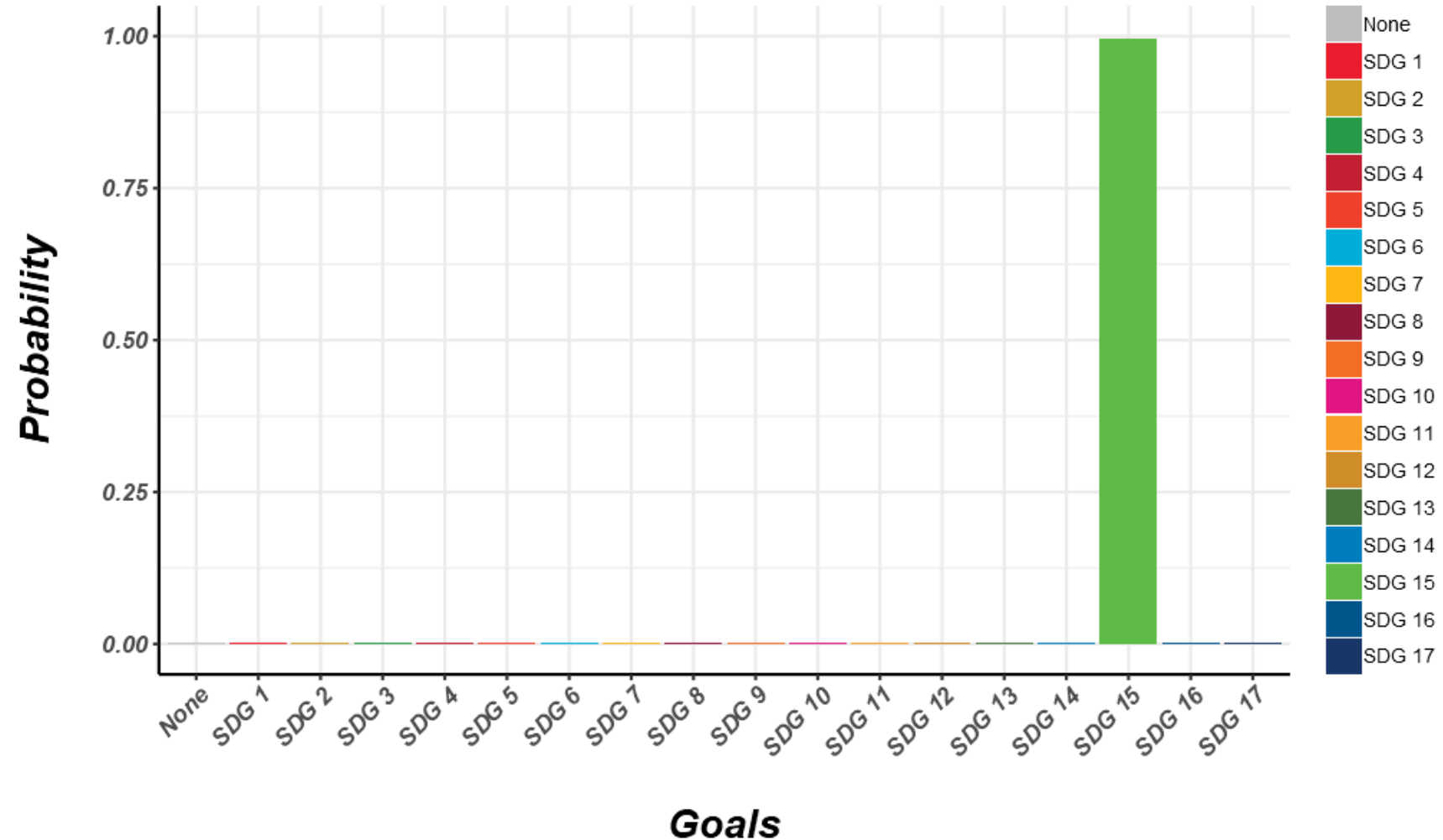
Extracting 'hidden' features

- Categorise by pooling these features and maximising the retention of features discriminating among categories
- Features: sequences of words
- Future: character-level sequence features (1million+ text)



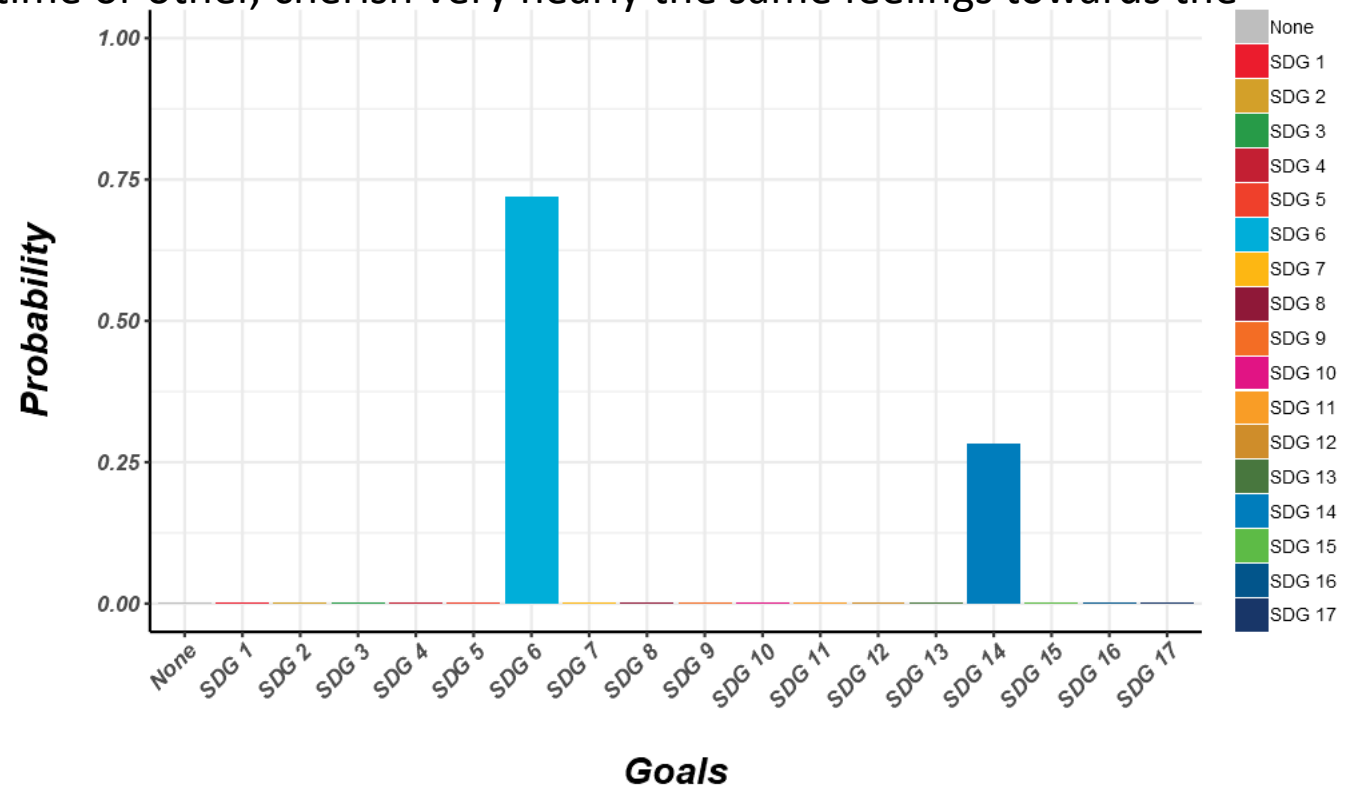
Trained model: prediction

“forest diversity is degraded by habitat loss”



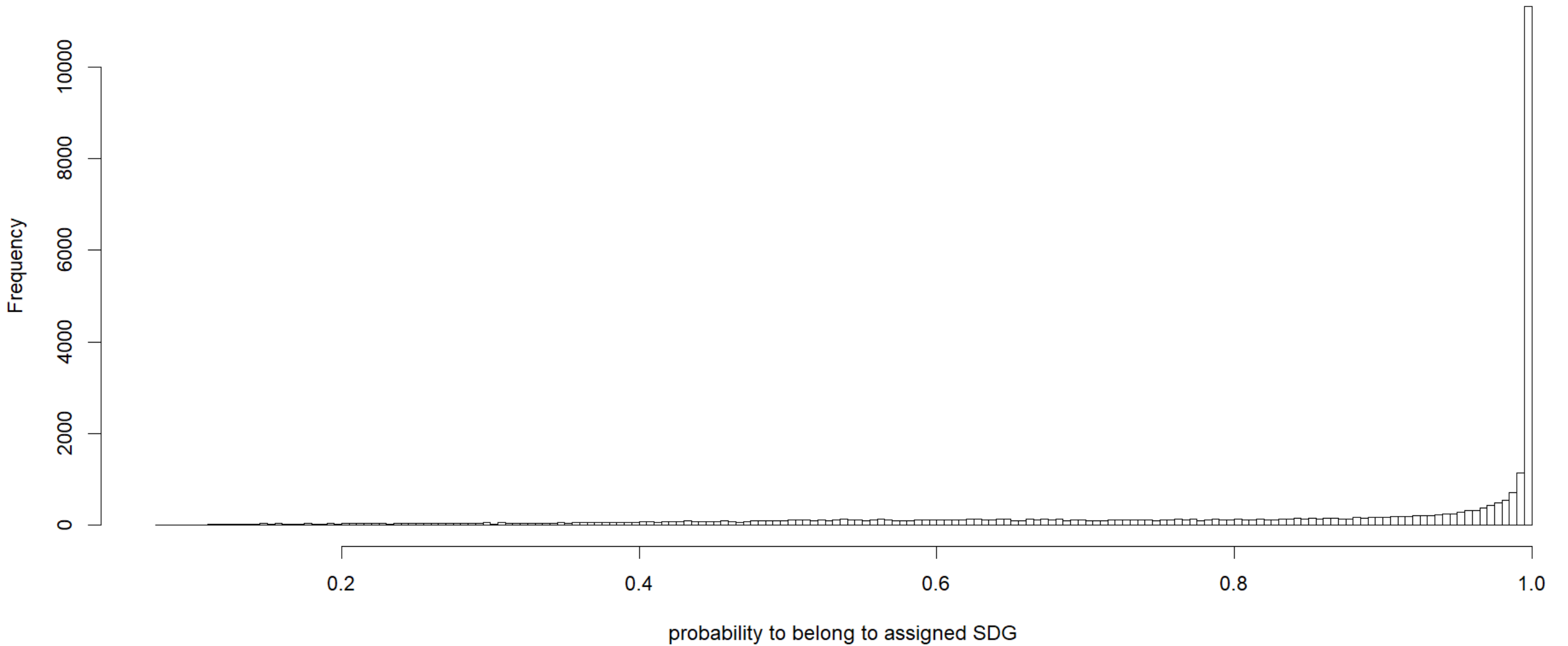
Trained model: prediction

“Call me Ishmael. Some years ago - never mind how long precisely - having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen and regulating the circulation. Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off - then, I account it high time to get to sea as soon as I can. This is my substitute for pistol and ball. With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship. There is nothing surprising in this. If they but knew it, almost all men in their degree, some time or other, cherish very nearly the same feelings towards the ocean with me.”



UoA – probability of max SDG

~32k outputs, ~30sec on laptop

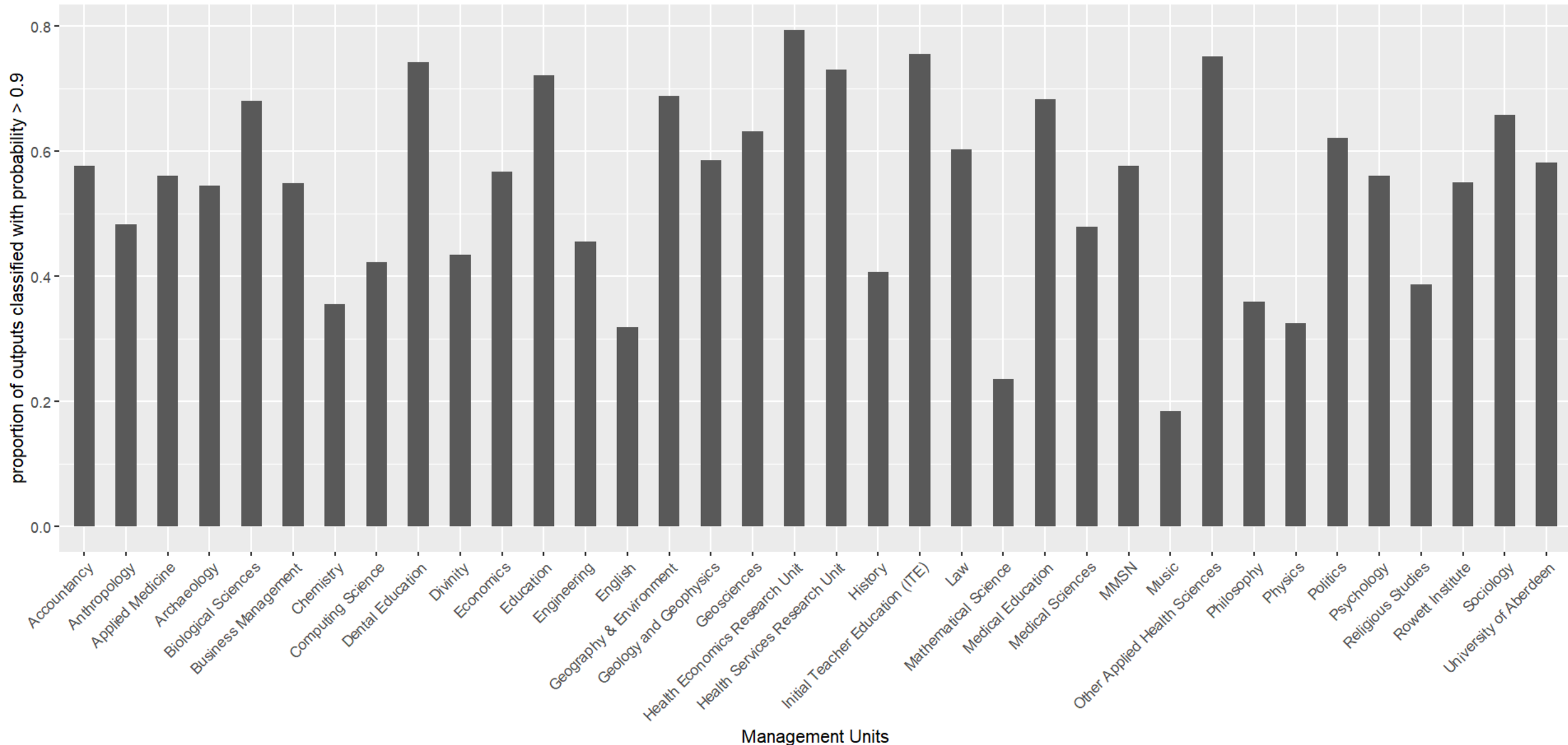


Inference: CNN predictions become observations

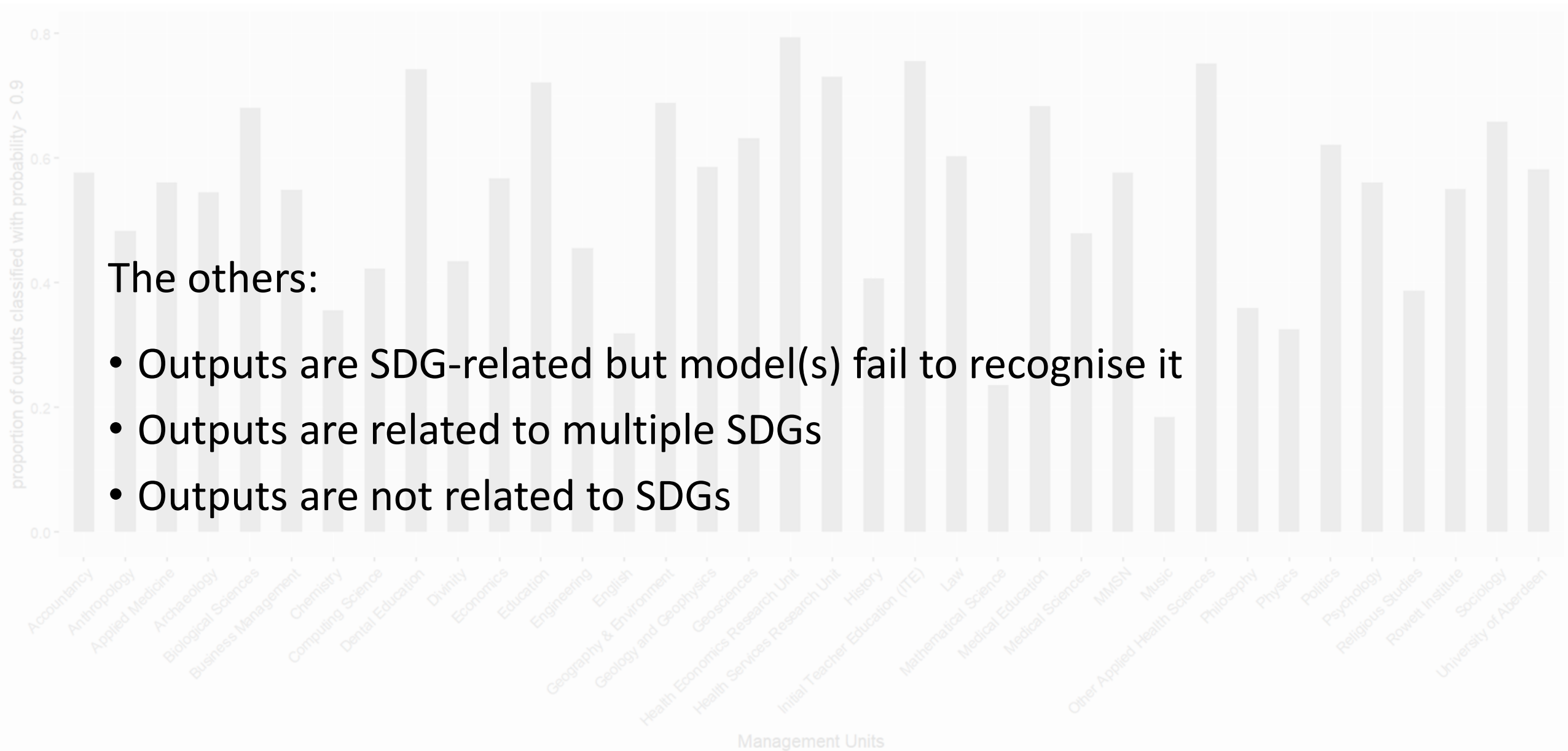
- What is the SDG landscape of a text corpus
- I would not have confidence to use it for single output assessment
- I would not have confidence to use it for sets of outputs with lower sample size

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
41000232	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41000284	0	0	16	5	2	1	0	0	0	0	3	0	0	0	0	5	0
41000954	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	2	0
41000969	0	1	0	2	1	1	0	0	0	0	0	1	0	0	0	0	0
41000970	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
41001072	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
41001100	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0
41001131	2	5	0	6	0	5	2	2	0	0	2	0	0	1	10	0	0
41001157	0	0	1	3	0	0	1	0	0	0	0	0	0	0	0	0	0
41001168	0	0	2	2	0	0	1	0	0	0	0	0	0	0	0	4	0
41001172	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
41001213	0	1	4	2	2	2	0	6	0	0	6	0	0	1	0	11	0
41001215	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0
41001222	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
41001239	0	7	0	5	1	6	3	2	0	0	4	1	0	2	6	2	0
41001240	0	0	0	0	0	0	3	1	0	0	1	0	0	0	0	0	0
41001300	0	4	0	3	2	0	0	1	0	0	1	0	0	1	7	1	0
41001309	0	1	5	4	0	0	2	1	0	0	3	0	0	0	0	1	0
41001323	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	3	0
41001335	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

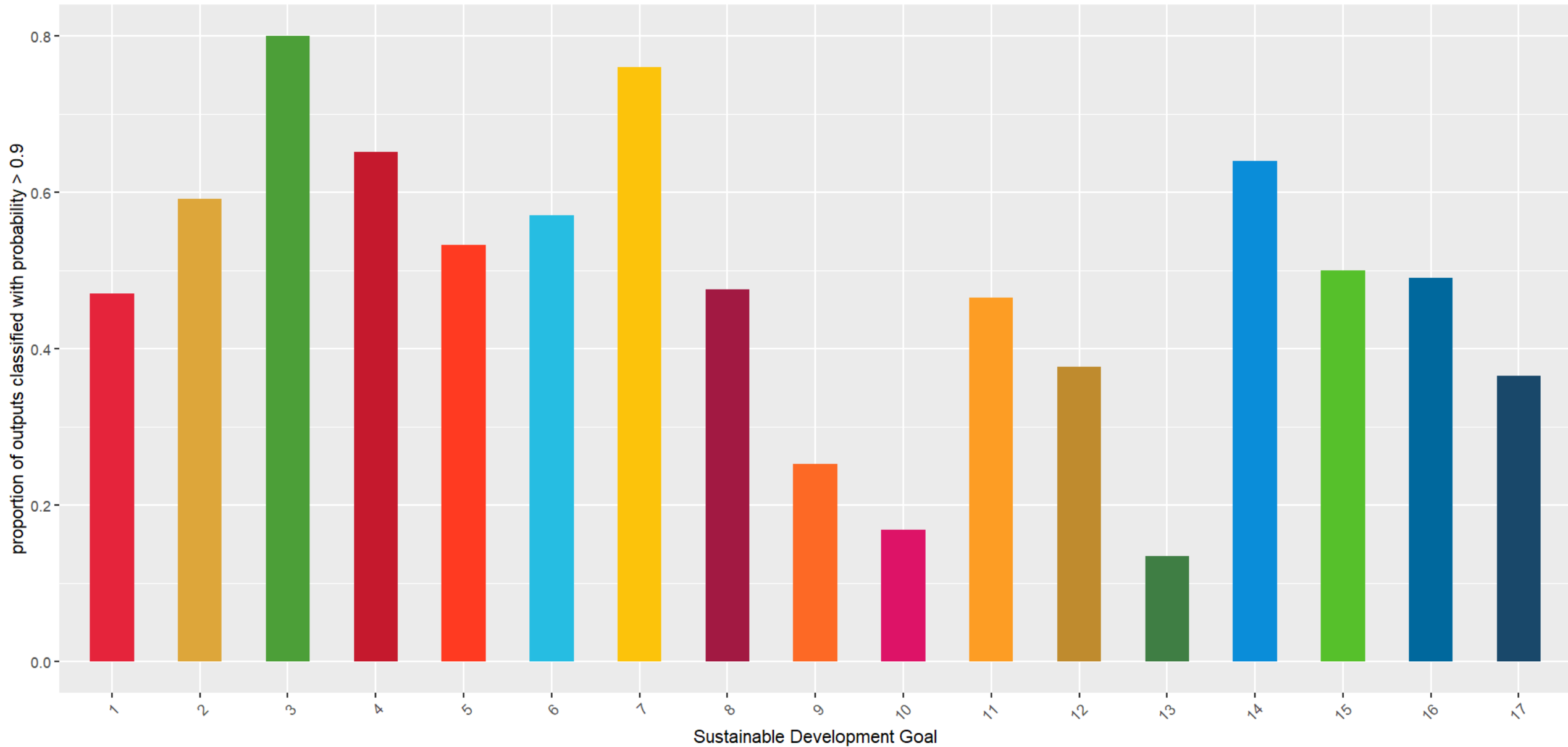
Abstracts classified confidently by MUs



Abstracts classified confidently by MUs

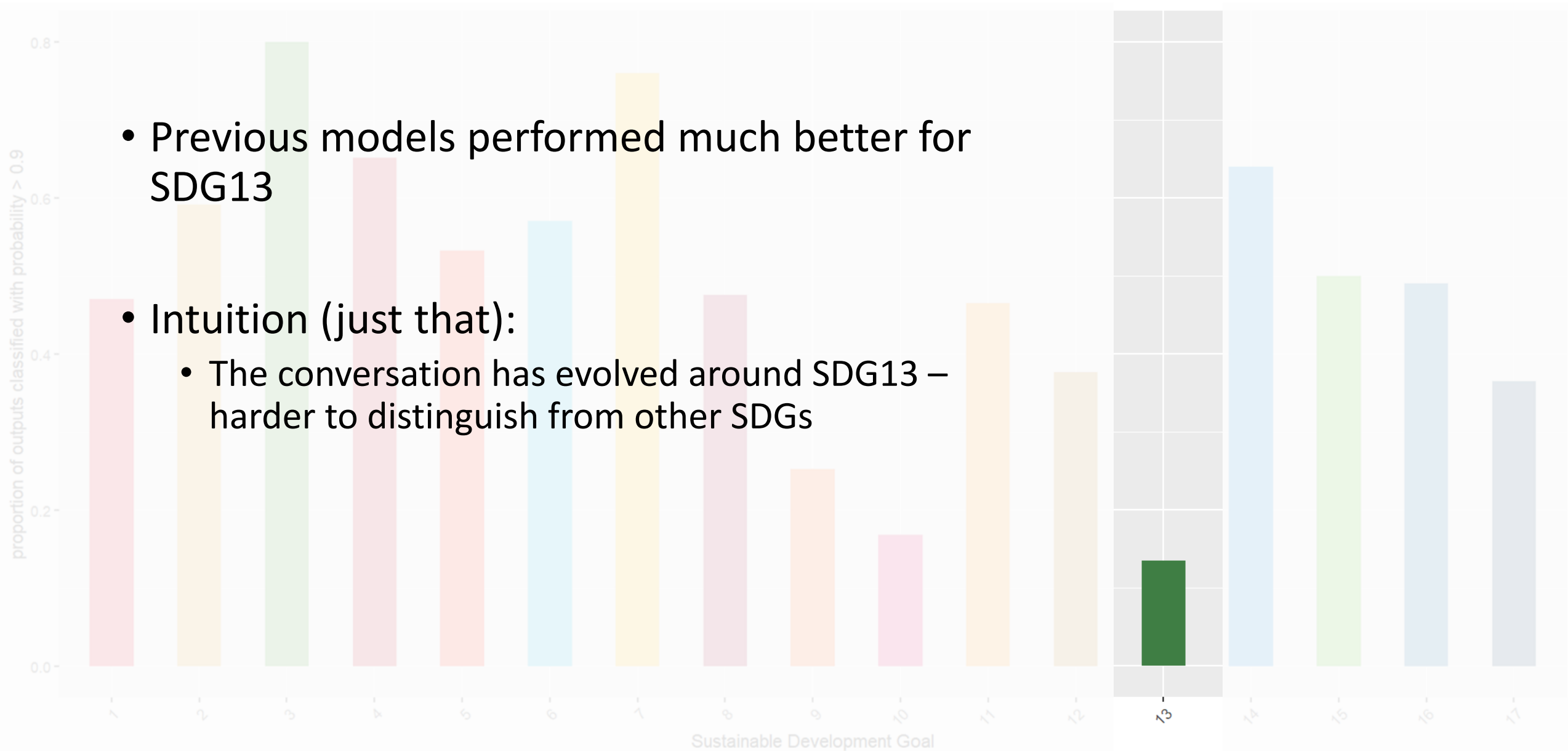


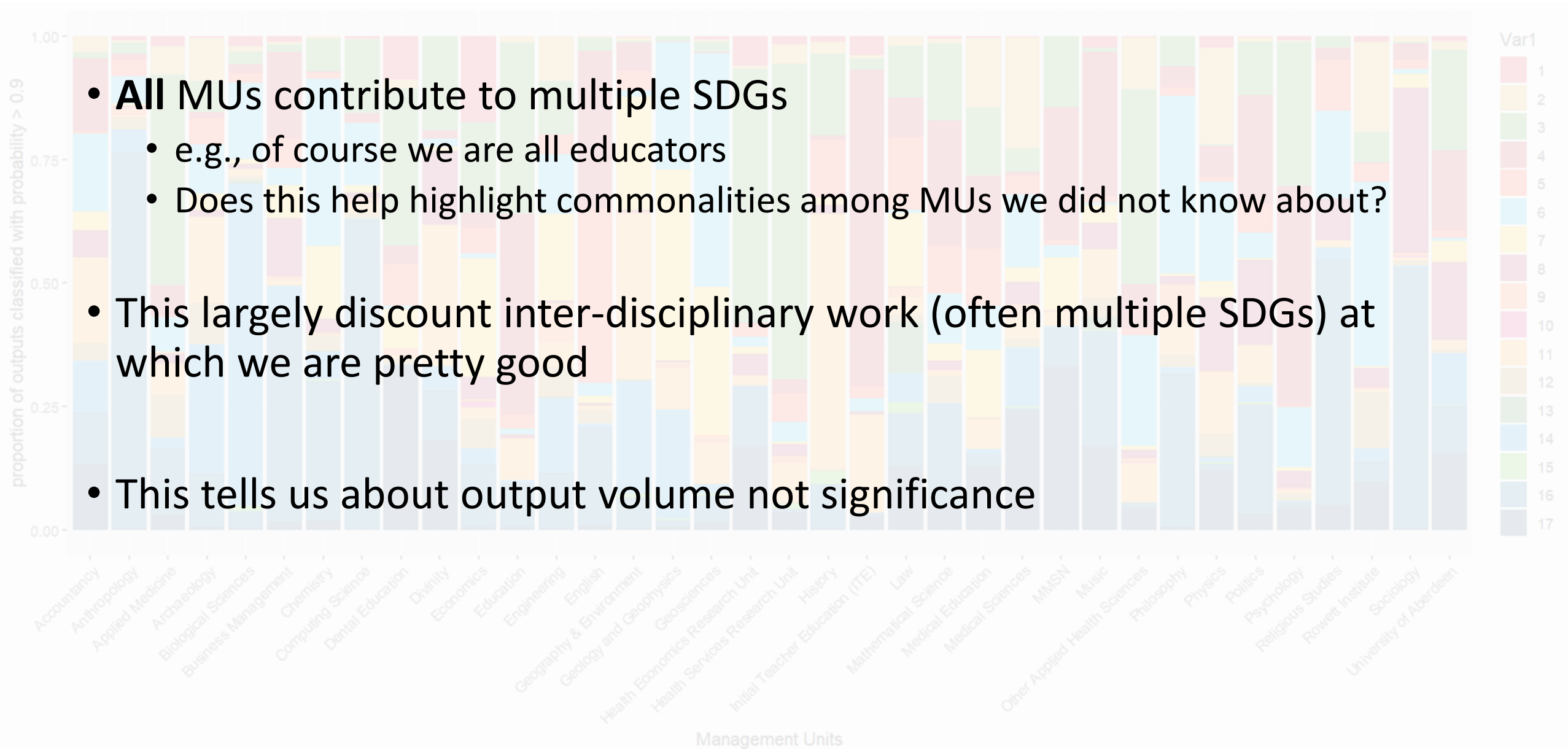
Abstracts classified confidently by SDGs



Why?

- Previous models performed much better for SDG13
- Intuition (just that):
 - The conversation has evolved around SDG13 – harder to distinguish from other SDGs





UoA SDG Landscape – high confidence outputs

