**Week 6 Update**

**Summary of thesis**

What this is telling us: some slides that show as high predictors are in later modules. Hard to say where in the module they occur. Slide importance might not be best for dropout prediction. There are some early slides that are predictive. Vanishing value.

Can calculate information gain between each feature and class variable – using IG for feature selection. Mutual information – there are other feature selection methods that RANK the features

Coefficient is not that widely used. Use a correlation based feature selection – find subset of features that are most important in predicting the class. Traditionally, logistic regression means between 0 and 1 – so try different classifiers – try decision trees – if DT is small, results are interpretable. Set of rules that useful for teachers – actually I can put this in RECOMMENDATIONS FOR FURTHER RESEARCH

Here are modules, here are slides, weigh them in diff. features – and take them to course designers. Which of these – do you find this useful, and score it. That gives you some external validation of your method – it’s impact and value. Think about what application this would have value in – can help content designers – automated system – getting educational value

Apply this method to uni course later on

**Generate encoding for one module and send to Irena**

This first experiment which used slide interaction sequences to predict final grade recorded an accuracy of 79% compared to 78% baseline. In the first figure, we can see that in predicting final grade, a portion of slides (to be quantified) are more correlated with higher final grade performance (e.g. slides that record a feature importance over 1.25), while other slides have weaker than average (1.00) correlation with final grade performance. The pedagogical value of this result is that educators can focus on a subset of slides that tend to be more correlated with higher grade performance, and can create interventions that encourage students to focus on those slides. It’s not actually the slide that impacts this – if you change score distribution so that you have more equal bins, then you will reduce the impact, will change from 75% to 25%.

Right now, we’re predicting final outcome – grade/dropout. Not as useful as predicting grade for this module/next module – or dropout across the next week. Utility – we want to care about if they’re going to drop out next week – because you have modules x students.

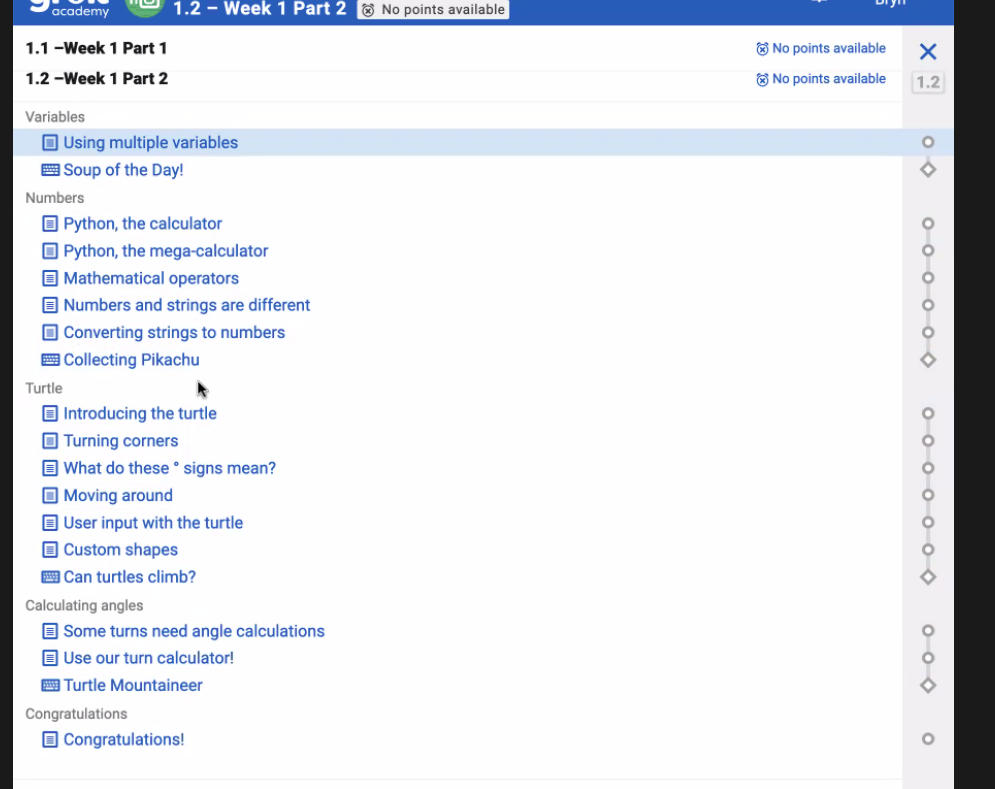
Predicting score might not be that useful, so focus on the dropout. We define dropout as not participating at all in the following module. If we just looked at the slides, and then predicting outcomes in that one module – does it pass on time (don’t allow late submissions), failed but didn’t pass, or didn’t attempt.

Are the slides helpful for the outcome of the module? Is the activity of that module indicative of the outcome.

Quantify value of slides in a module – predict one

For each module there are 4 problem slides

Try writing this as a precursor paper -



We could just look at the final problem in each module – and predict that – final problem is a capstone for the material

Pair problems – first problem has more scaffolding, second problem has less scaffolding – what’s the value of the material – was the slide used or not used, and what was its effect

Show that these slides are useful – and then show how that compares with students completing it or not

PHASE 2 refinement: Next thing to do is did they do anything on that slide or problem, and then use that to predict it (dropping the bar in terms of student effort – not only passed the interactive slide). Likely to be that earlier problems are the most predictive.

Slide improvement: 0 no attempt, 1 slide steps attempt, 2 slide steps complete

Take all slides – interactive slides and problem slides

Do for each module separately

**Outcomes**

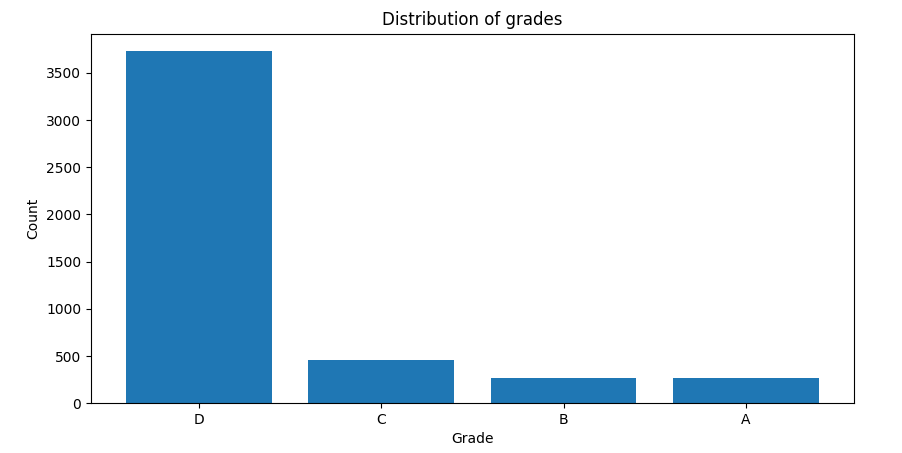
* **Defined two types of sequences:** slide interaction sequences containing only slides and problem interaction sequences containing only problems.
* **Grade prediction:** discretised scores into four grade categories, and predicted student grades using number of slides completed and slide interaction and problem interaction sequences.
* **Cross-fold validation with stratification:** validated all prediction results with 10-fold cross validation with stratification (do we need to show performance for all folds?)
* **Used slide interaction sequences and problem interaction sequences** to predict dropout and grade.
* **Conducted feature importance analysis** to analyse which slides were most predictive of no dropout and high grade attainment.

**Grade prediction**

Last time, we saw that predicting final challenge scores for students was quite difficult, with low accuracy rates observed. In part, this was because of the right-skewed distribution of scores, with several students recording 0s (i.e. no problems passed).

To partly remedy this problem, I created four grade categories, and discretised the scores into those grade categories. The current discretisation method takes **strict quartiles** of score ranges e.g. 0-99 is a D, 100-199 is a C, 200-299 is a B, 300-400 is an A. It may be worth **exploring other discretisation methods** to create a more even distribution than below.

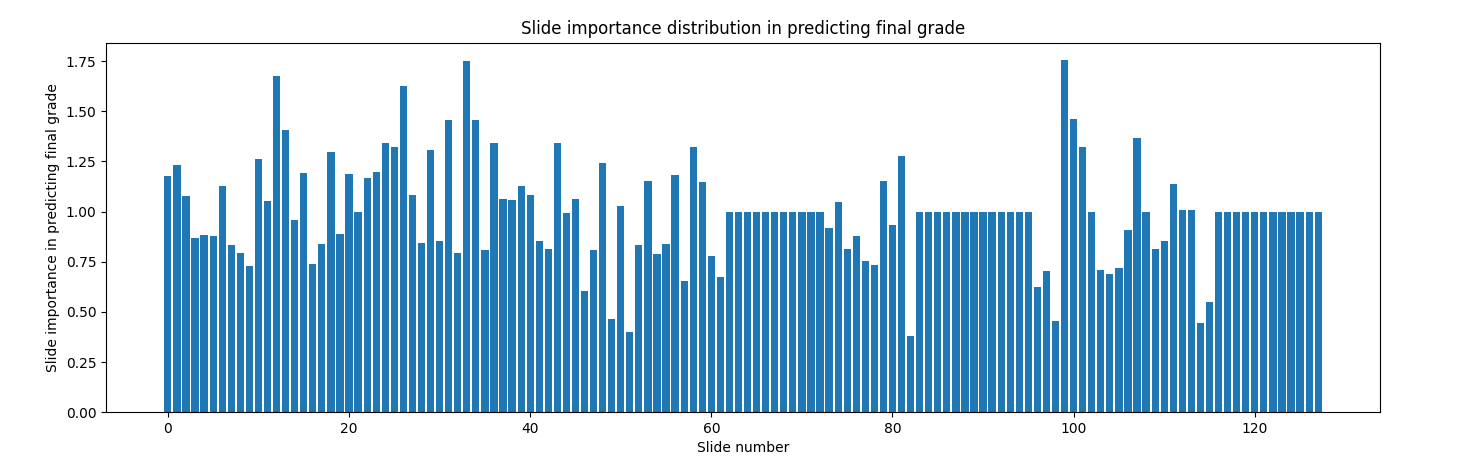
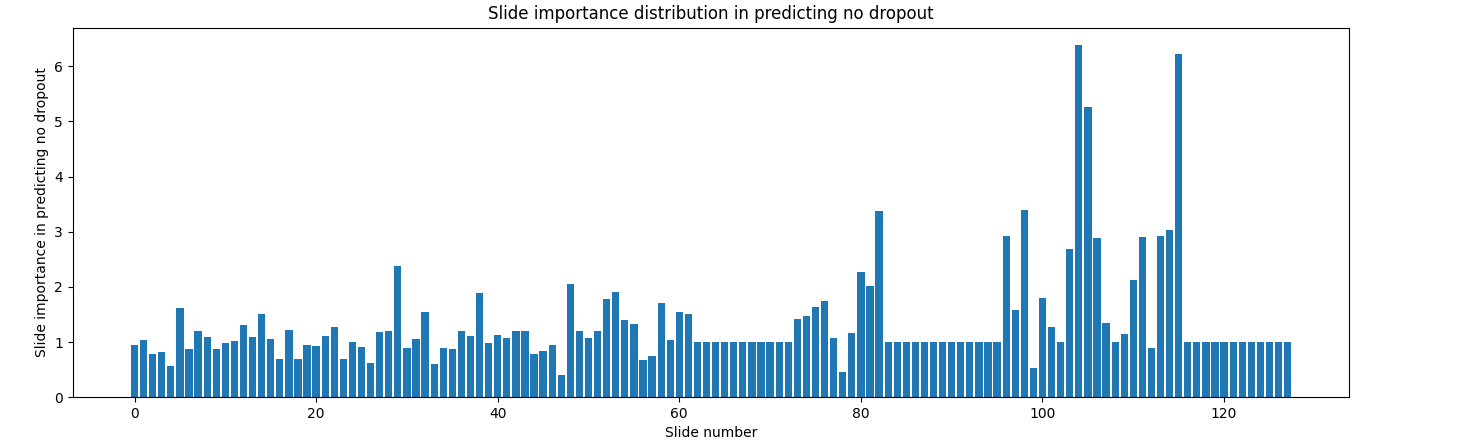
What proportion is 0 – if it’s a separate value, treat it as its own. Then you can go across distribution to make sure it’s more evenly spread. Do quartiles of the spread of population. Check with Sophia’s discretisation method. She used 0-100 including, 201 – 300. Then, we would have better binning overall.



**Grade prediction**

This week, I also conducted feature importance by looking at the regression coefficients for logistic regression, and plotting the feature (slide) importance across all 120 slides in the 2018-newbies-challenge in predicting 1) final grade and 2) dropout. The features used are the slide interaction binary sequences, indicating whether the slides were completed or not. In a second pair of experiments, the features used are the problem interaction binary sequences, indicating whether the problems were completed or not.

**Note on feature importance:** The logistic regression coefficient β associated with a predictor X is the expected change in log odds of having the outcome per unit change in X. So increasing the predictor by 1 unit (or going from 1 level to the next) multiplies the odds of having the outcome by eβ.



What this is telling us: some slides that show as high predictors are in later modules. Hard to say where in the module they occur. Slide importance might not be best for dropout prediction. There are some early slides that are predictive. Vanishing value.

Can calculate information gain between each feature and class variable – using IG for feature selection. Mutual information – there are other feature selection methods that RANK the features

Coefficient is not that widely used. Use a correlation based feature selection – find subset of features that are most important in predicting the class. Traditionally, logistic regression means between 0 and 1 – so try different classifiers – try decision trees – if DT is small, results are interpretable. Set of rules that useful for teachers – actually I can put this in RECOMMENDATIONS FOR FURTHER RESEARCH

Here are modules, here are slides, weigh them in diff. features – and take them to course designers. Which of these – do you find this useful, and score it. That gives you some external validation of your method – it’s impact and value. Think about what application this would have value in – can help content designers – automated system – getting educational value

Apply this method to uni course later on

**Generate encoding for one module and send to Irena**

This first experiment which used slide interaction sequences to predict final grade recorded an accuracy of 79% compared to 78% baseline. In the first figure, we can see that in predicting final grade, a portion of slides (to be quantified) are more correlated with higher final grade performance (e.g. slides that record a feature importance over 1.25), while other slides have weaker than average (1.00) correlation with final grade performance. The pedagogical value of this result is that educators can focus on a subset of slides that tend to be more correlated with higher grade performance, and can create interventions that encourage students to focus on those slides. It’s not actually the slide that impacts this – if you change score distribution so that you have more equal bins, then you will reduce the impact, will change from 75% to 25%.

Right now, we’re predicting final outcome – grade/dropout. Not as useful as predicting grade for this module/next module – or dropout across the next week. Utility – we want to care about if they’re going to drop out next week – because you have modules x students.

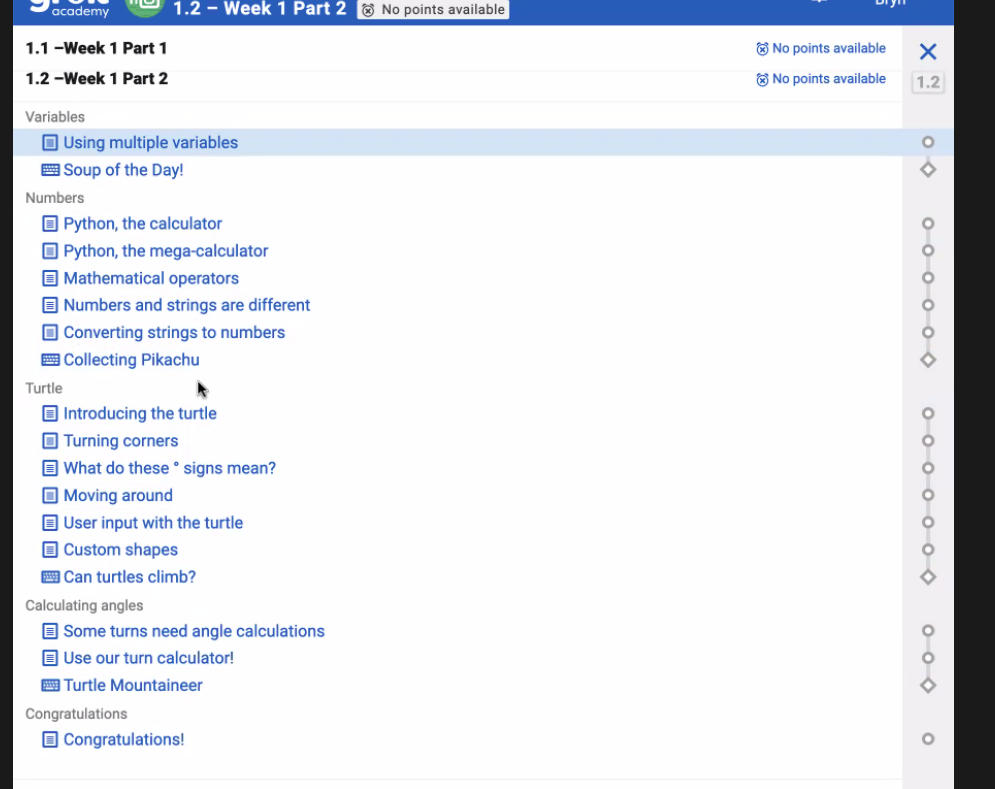
Predicting score might not be that useful, so focus on the dropout. We define dropout as not participating at all in the following module. If we just looked at the slides, and then predicting outcomes in that one module – does it pass on time (don’t allow late submissions), failed but didn’t pass, or didn’t attempt.

Are the slides helpful for the outcome of the module? Is the activity of that module indicative of the outcome.

Quantify value of slides in a module – predict one

For each module there are 4 problem slides

Try writing this as a precursor paper -



We could just look at the final problem in each module – and predict that – final problem is a capstone for the material

Pair problems – first problem has more scaffolding, second problem has less scaffolding – what’s the value of the material – was the slide used or not used, and what was its effect

Show that these slides are useful – and then show how that compares with students completing it or not

PHASE 2 refinement: Next thing to do is did they do anything on that slide or problem, and then use that to predict it (dropping the bar in terms of student effort – not only passed the interactive slide). Likely to be that earlier problems are the most predictive.

Slide improvement: 0 no attempt, 1 slide steps attempt, 2 slide steps complete

Take all slides – interactive slides and problem slides

Do for each module separately

This second experiment which used slide interaction sequences to predict dropout recorded an accuracy of 84% compared to 77% baseline. In the second figure, we can see like above, there are a few slides that are more strongly correlated with students not dropping out than others, particularly slides towards the latter stages of the challenge (e.g. slide 100+), as well as a few slides in the slide number range of 80-100. This demonstrates that students who tend to not dropout of the course i.e. complete all problems, typically are able to successfully complete the latter stages of the challenge. .

In this third experiment, we use problem interaction sequences to predict dropout grade, recording an accuracy of 99% compared to a baseline of 77%. We then use half of the problem interaction sequence to predict dropout, and record an accuracy of 91% compared to a baseline of 78%. This demonstrates that even by simply using half of the history of a student’s completion of problems, we can predict whether or not they will dropout at the end of the course with 99% accuracy (this has validated with cross-validation).