

# Keep-Me-Engaged

Customized Recommendations to Improve e-Learner Engagement

*Lindsay Warrenburg*



# New Tool for Premium Subscribers: *Customized Recommendations*

The Zoom logo, consisting of the word "zoom" in a blue, lowercase, sans-serif font.The GoToTraining logo, featuring a purple flower-like icon to the left of the text "GoToTraining" in a black, sans-serif font.The Cisco Webex logo, featuring a blue and green circular icon to the left of the text "Cisco webex" in a green, lowercase, sans-serif font.The Segment logo, featuring a green circular icon to the left of the word "Segment" in a green, lowercase, sans-serif font.The Stripe logo, consisting of the word "stripe" in a blue, lowercase, sans-serif font.The Zapier logo, featuring a red sun-like icon above the word "zapier" in a red, lowercase, sans-serif font.The Salesforce logo, featuring a blue cloud-like shape with the word "salesforce" in white, lowercase, sans-serif font.The Adobe Connect logo, featuring a green square icon with a white cross-like shape to the left of the text "Adobe Connect" in a black, sans-serif font.The Insight logo, featuring the word "INSIGHT" in a black, uppercase, sans-serif font.

# New Tool for Premium Subscribers: *Customized Recommendations*

7,000+  
Companies



200,000+  
Courses



Amazon **Redshift**



Postgre**SQL**

**Access and Download Data**  
*9.9B data points*



Amazon **Redshift**

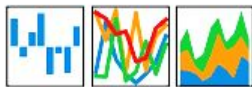


Postgre**SQL**

**Access and Download Data**  
*9.9B data points*



python



$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

pandas



**Feature Engineering**  
**Random Forest Regression**



Amazon Redshift

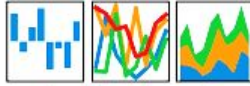


PostgreSQL

**Access and Download Data**  
*9.9B data points*



python



$$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$$

pandas



**Feature Engineering**  
**Random Forest Regression**



Streamlit



HEROKU

**Web App Creation**



# Learner Engagement



## Learner Engagement



### Admin. Features:

Price of Course

Length of Course





## Learner Engagement



### Admin. Features:

Price of Course

Length of Course



### Learner Features:

Time Spent on Course

Average Quiz Grade



## Learner Engagement



### Admin. Features:

Price of Course  
Length of Course



### Instructor Features:

% of Non-Graded Quizzes  
Number of Collaborations



### Learner Features:

Time Spent on Course  
Average Quiz Grade



## Learner Engagement



### Admin. Features:

Price of Course  
Length of Course



### Instructor Features:

% of Non-Graded Quizzes  
Number of Collaborations



### Learner Features:

Time Spent on Course  
Average Quiz Grade



### Website Features:

% Interactive Content  
% Videos



## Learner Engagement



### Admin. Features:

Price of Course  
Length of Course



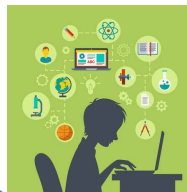
### Learner Features:

Time Spent on Course  
Average Quiz Grade



### Instructor Features:

% of Non-Graded Quizzes  
Number of Collaborations

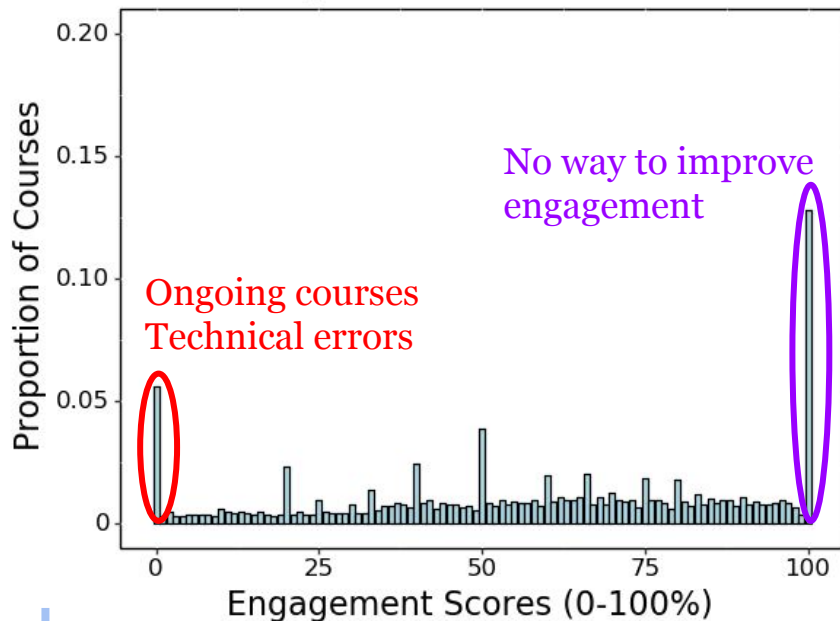


### Website Features:

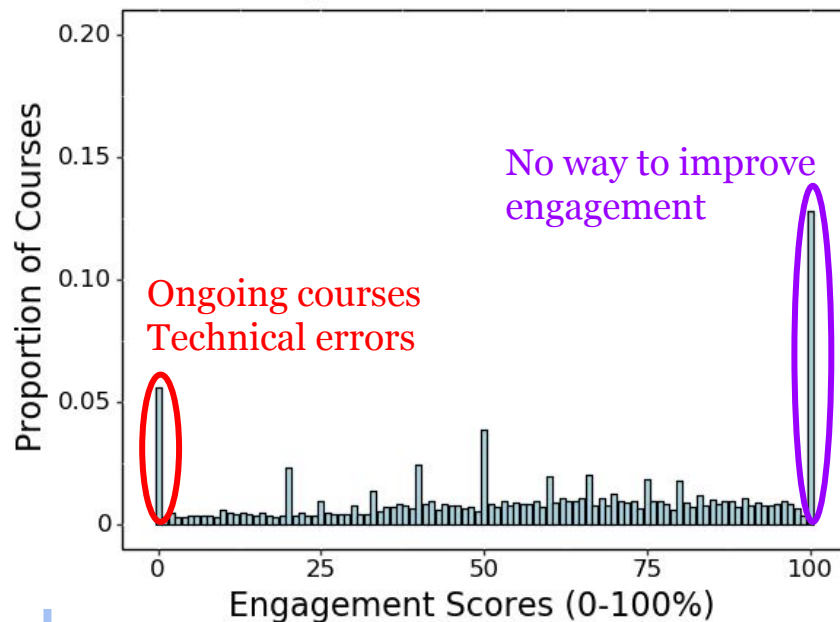
% Interactive Content  
% Videos

# The distribution of Engagement Scores is highly irregular

Original Distribution



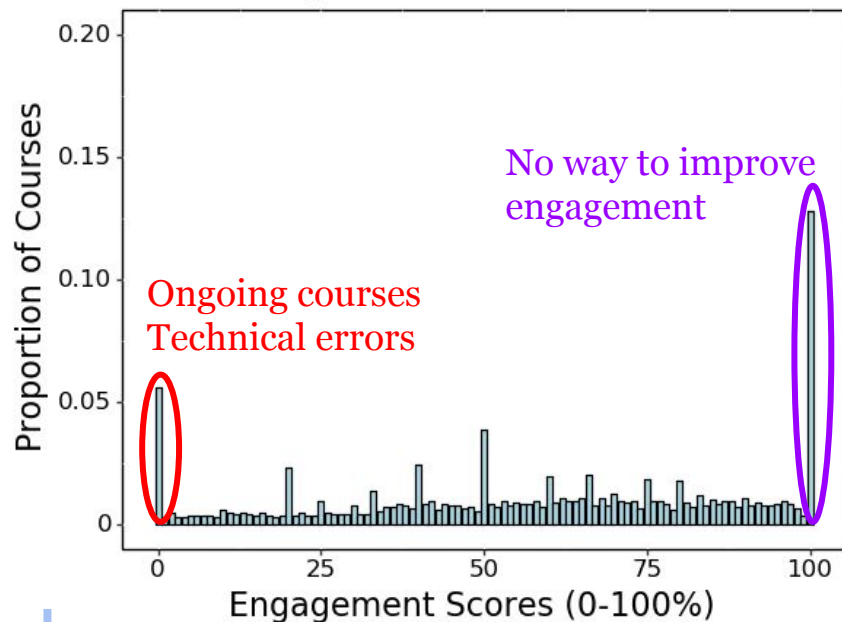
## Original Distribution



## Scope of Keep-Me-Engaged

- Help course developers with Engagement Scores of 1-99% understand current engagement scores

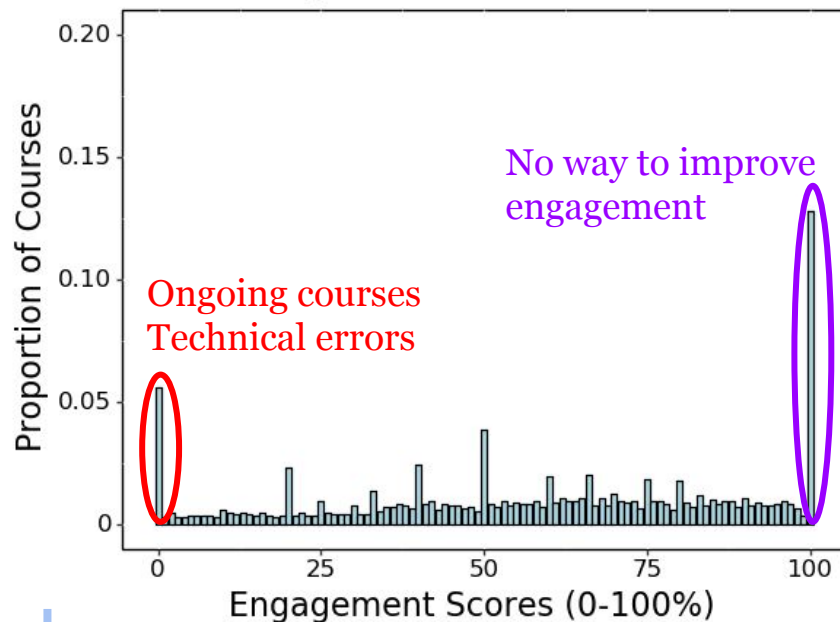
## Original Distribution



## Scope of Keep-Me-Engaged

- Help course developers with Engagement Scores of 1-99% understand current engagement scores
- Learn how to improve engagement scores for future iterations of the course (e.g., next semester)

## Original Distribution

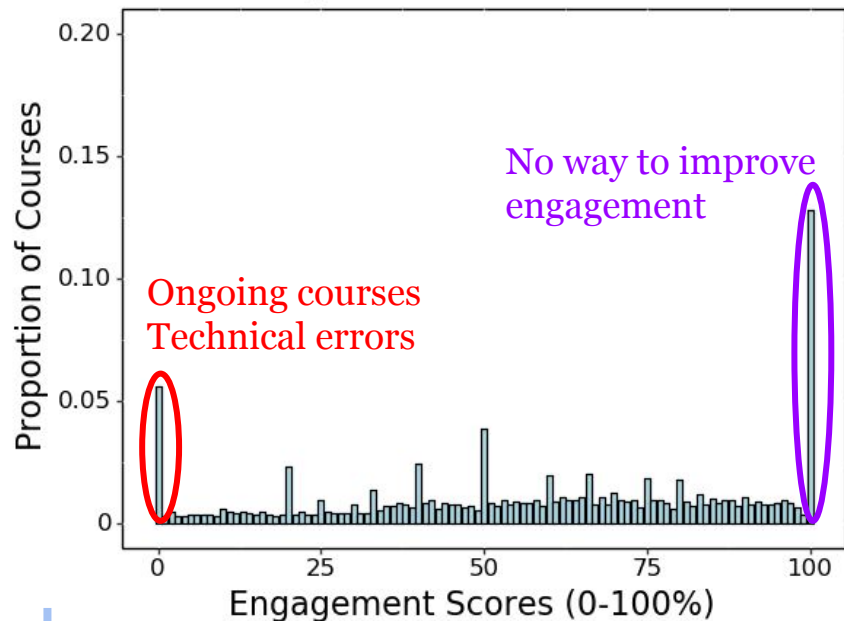


## Scope of Keep-Me-Engaged

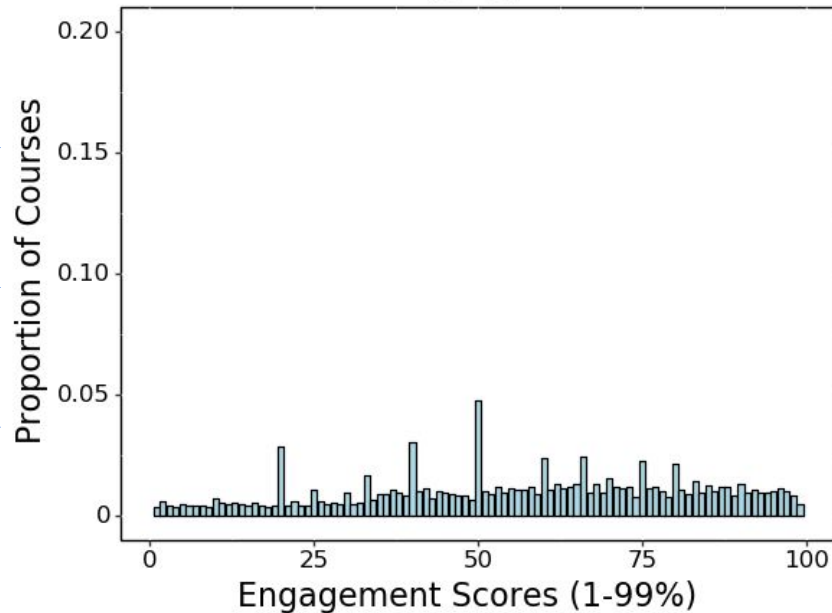
- Help course developers with Engagement Scores of 1-99% understand current engagement scores
- Learn how to improve engagement scores for future iterations of the course (e.g., next semester)
- *Not aimed at predicting engagement scores of a brand new course*



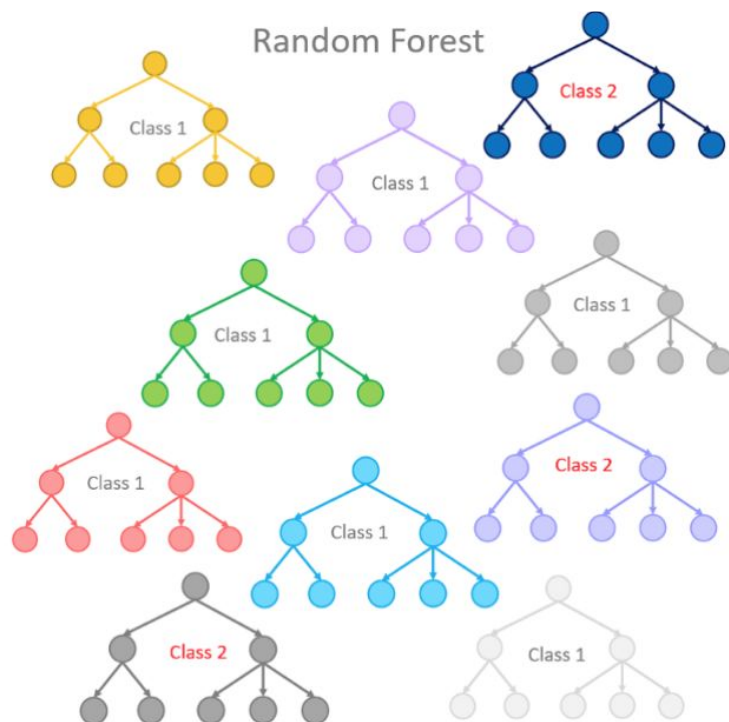
## Original Distribution



## Continuous Engagement Scores



# Predicting User Engagement Scores

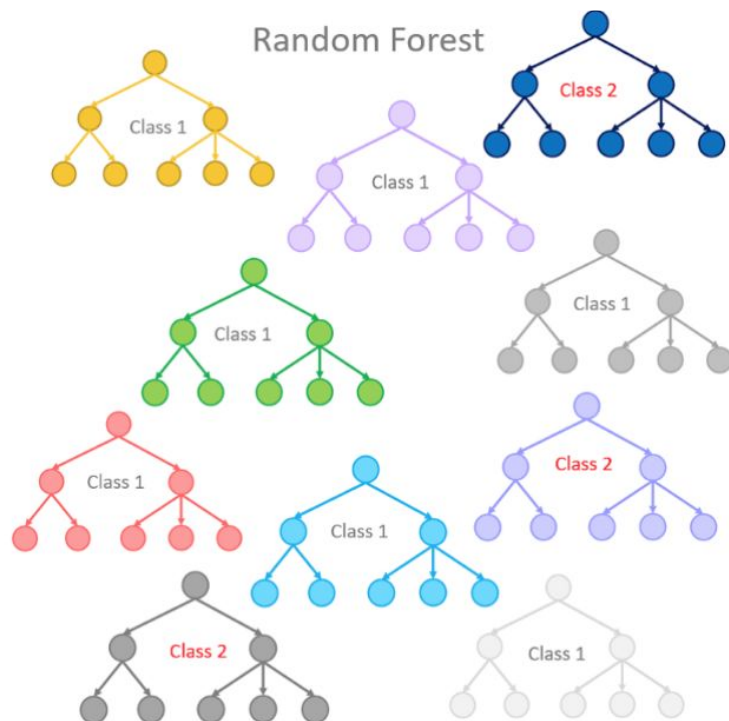


## Random Forest Regression

$$R^2 = 0.62$$

*Train:Test split = 80:20*

# Predicting User Engagement Scores



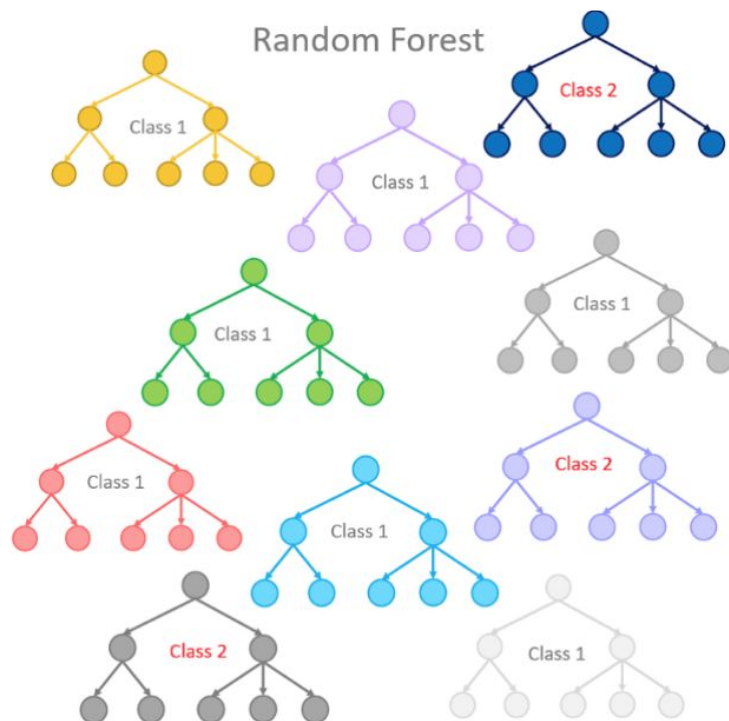
## Random Forest Regression

$$R^2 = 0.62$$

## Linear Regression (log)

$$R^2 = 0.16$$

# Predicting User Engagement Scores



## Random Forest Regression

$$R^2 = 0.62$$

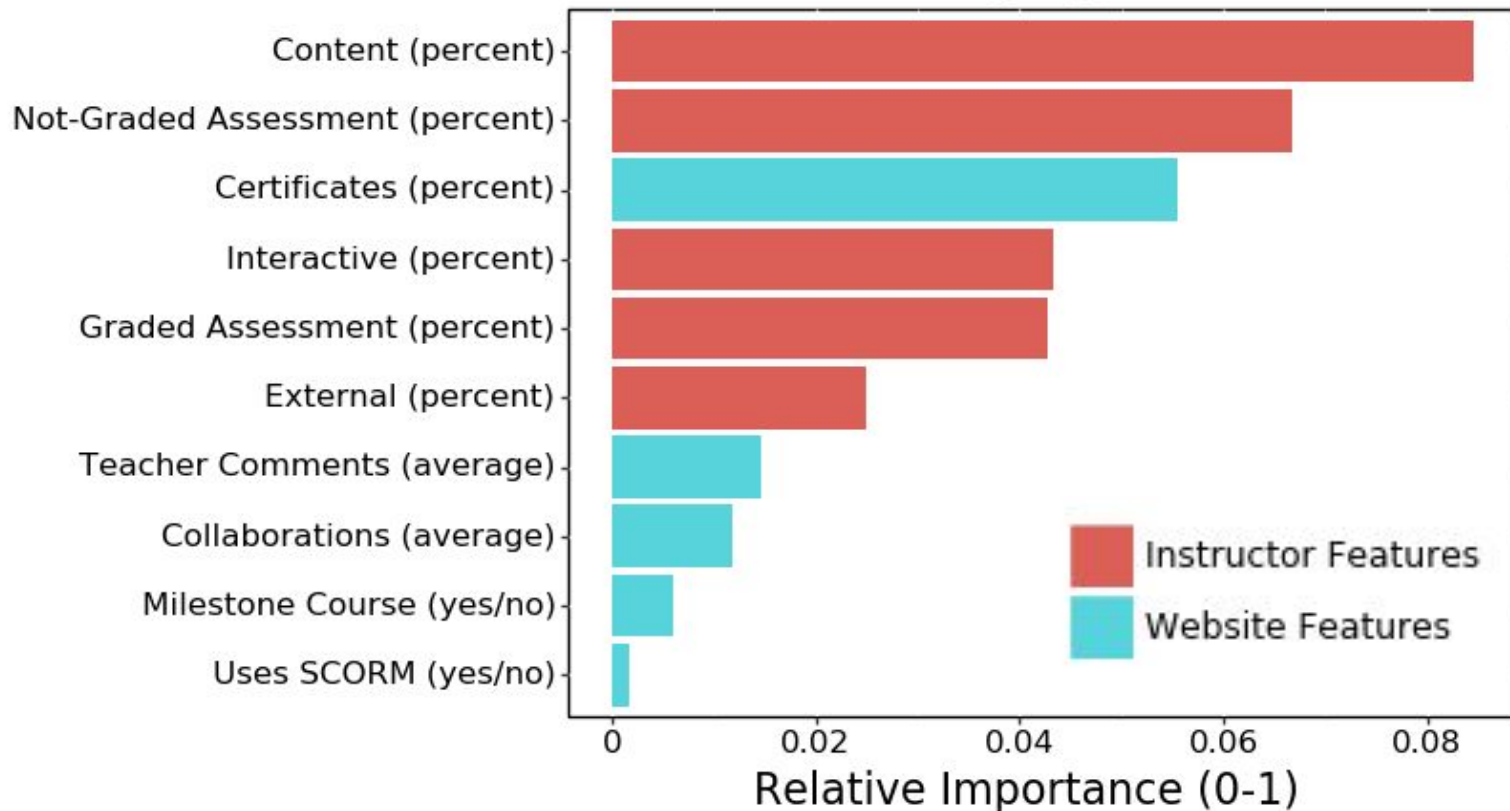
## Linear Regression (log)

$$R^2 = 0.16$$

## Ridge Regression (log)

$$R^2 = 0.16$$

# Feature Importance for User Engagement



# Interactive *Keep-Me-Engaged* Tool

<https://keep-me-engaged.herokuapp.com/>

# Thought Industries will use this feature as part of their Premium Package

Customized recommendations for

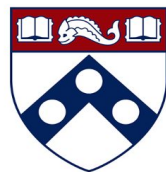
30,000+ courses

550+ companies

# Lindsay Warrenburg



Market Research • Data Visualization • B2B Data



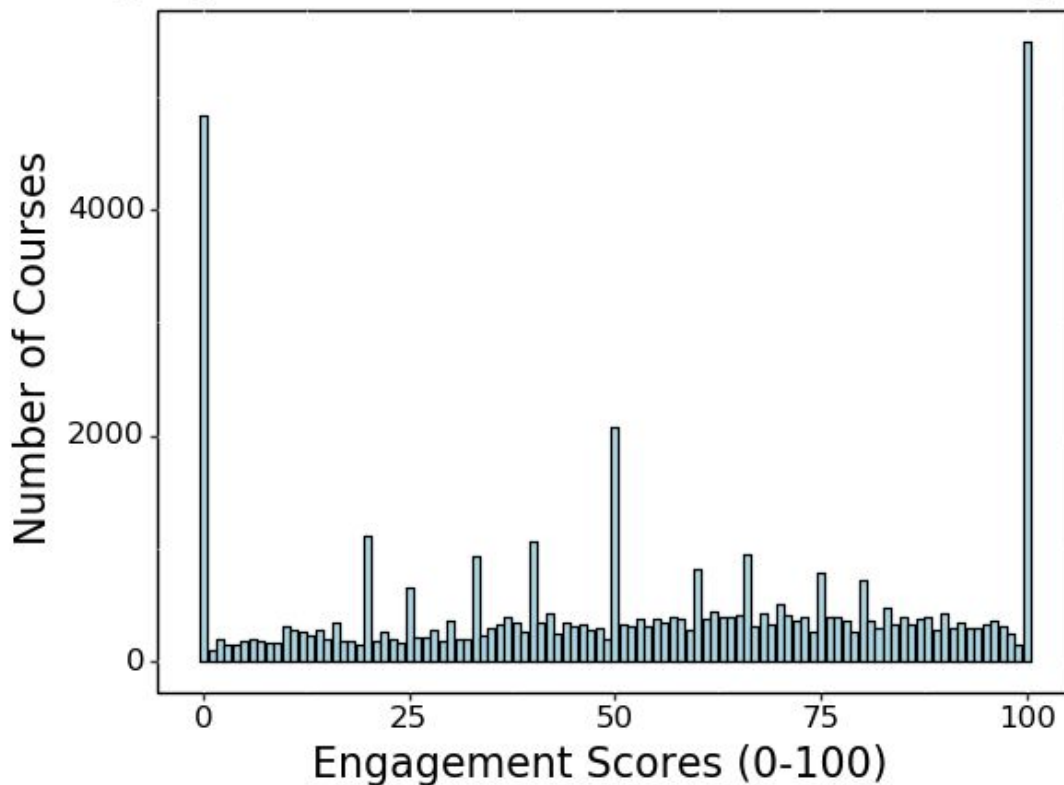
**Penn**  
UNIVERSITY of PENNSYLVANIA



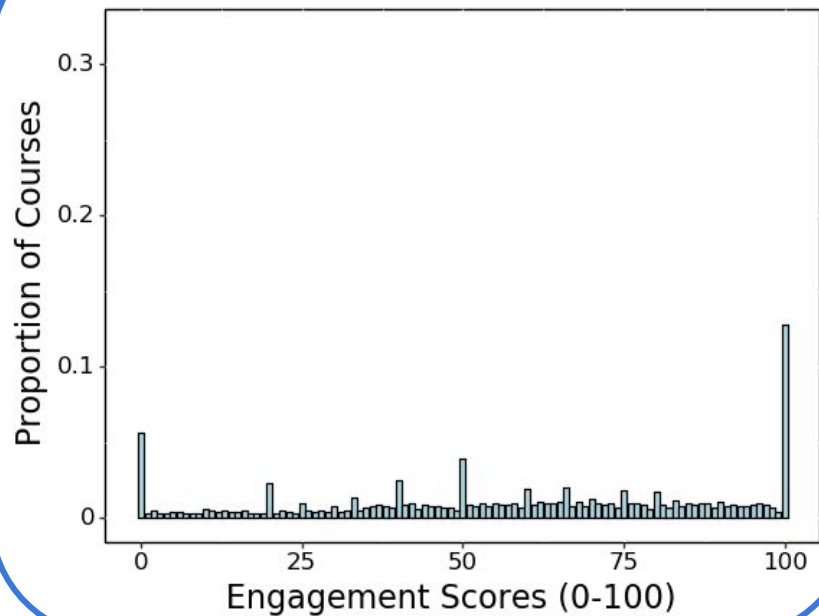
Music • Cognition & Perception • Emotion



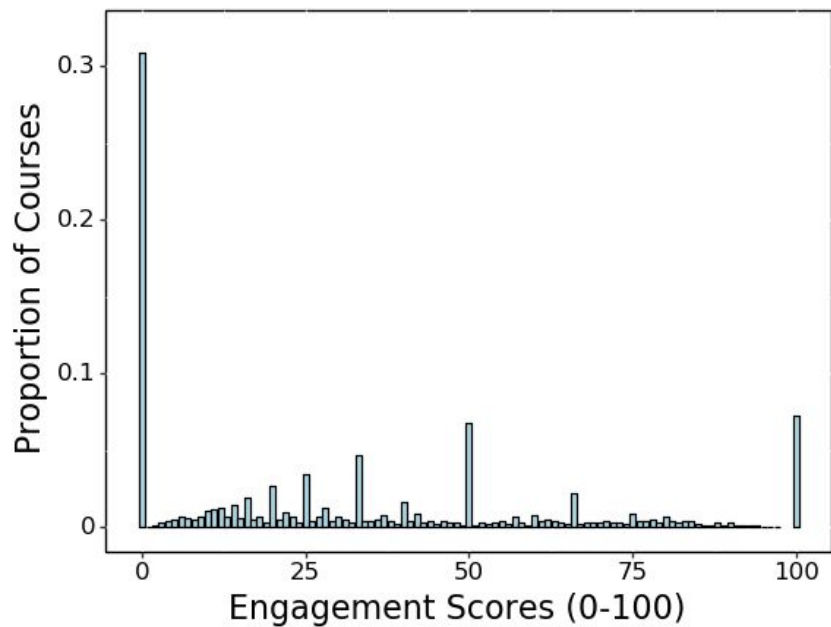
## Engagement Across All Course Types



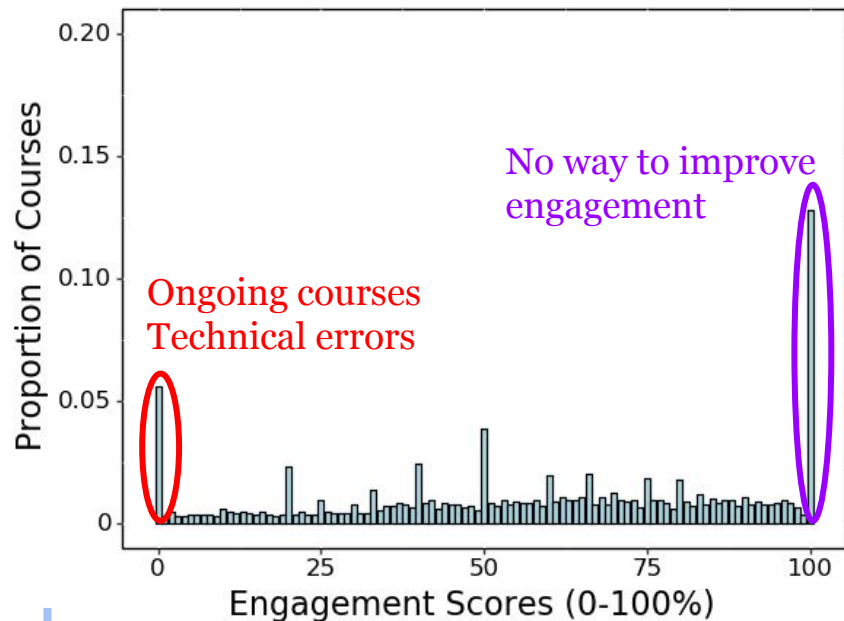
## Online Courses



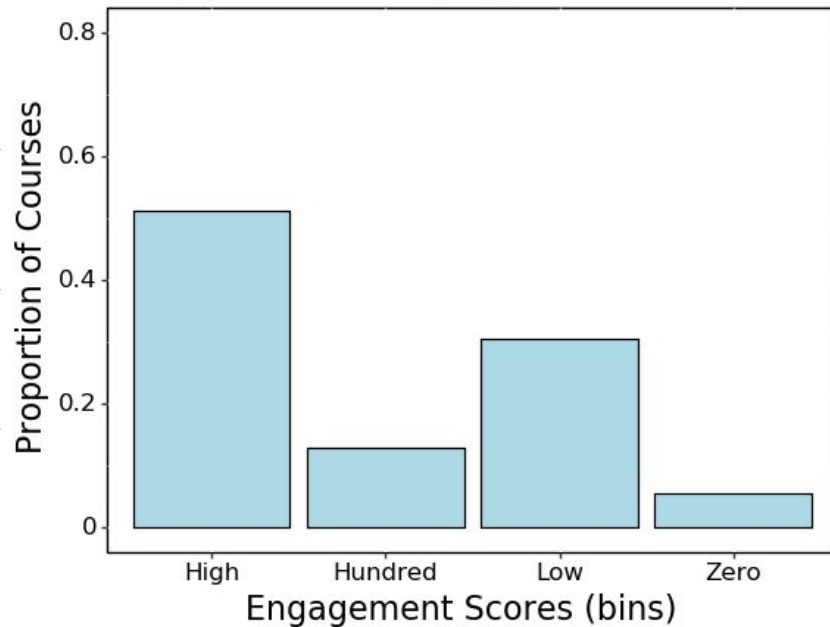
## Articles & Videos



## Original Distribution



## Categorical Engagement Scores



# Missing Values

- **Delete some courses with missing values**
  - *Example:* Courses that are missing our target (Engagement Scores)
- **Replace missing values with 0**
  - *Example:* No value for Number of Collaborations = no collaborations
- **Combine multiple features into one summary feature when only half the courses use SCORM at all**
  - *Example:* Average SCORM score, % SCORMs completed, % API SCORM, % Shareable SCORM ⇒ “Uses SCORM yes/no”
- **Fill in missing values with the median value**
  - *Example:* Average quiz grade

# Feature Importance for User Engagement

