

Keep-Me-Engaged

Customized Recommendations to Improve e-Learner Engagement

Lindsay Warrenburg



Interactive *Keep-Me-Engaged* Tool

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

Description of Feature Selection for the Thought Industries Team

Data Generation:

Selecting Top ~20 features from 602 possibilities

- **Examine summary data about each feature**
 - How many courses does this feature apply to?
If only 100 out of 200,000 courses use this feature, it isn't as helpful as a feature that 195,000 courses use
 - What kind of information is this?
Numerical data (Quiz Scores) can sometimes be more informative than text data (Quiz Passed: True/False)

Data Generation:

Selecting Top ~20 features from 602 possibilities

- **Prioritize features inside the project scope**

- We were interested in examining features that course instructors & course designers can use. Prioritize these features over eCommerce features.

- Write down whether that feature will help **measure** engagement or **predict** engagement

*Whether or not a user completed the course **measures** engagement*

*Whether or not certificates are offered could **predict** engagement*

Data Generation: Selecting Top ~20 features from 602 possibilities

- **Summarize information across all users**
 - Summarize all course information into one row in an Excel spreadsheet
Average quiz score across all learners in that course
 - If there's information that can't be summarized, don't include it
The email addresses of every course learner shouldn't affect engagement scores

Feature Selection & Transformation

- **Standardize measurement units**
 - Try to make the units of measurement similar across all variables
“Percent of Courses that Use this feature”
- **Continue to limit scope**
 - Delete courses that are “pending” or “drafts”
 - Only look at courses that have been viewed at least 1 time
- **Delete features that are highly correlated with Engagement Scores**
 - Example: Average number of hours spent on the course

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