Virtual Internship Data - ADS2001

Virtual internships are educational simulations in which students act as interns at a fictional company. The goal of virtual internships is to help students learn the ways of thinking, acting, talking, and justifying associated with a particular profession. In the virtual internship *Nephrotex*, students intern at a biomedial engineering company. During the internship, they work together on teams to design a prototype device to assistant patients with kidney failure. Their tasks include conducting background research, understanding stakeholder needs, designing prototypes, testing and evaluating prototypes, and justiying their design decisions. Teams work together in an online environment and communicate with one another via a chat tool. Teams can also communicate with a mentor assigned to their team who is there to ask questions and lead reflective discussions. For more information about virtual internships see: https://asmedigitalcollection.asme.org/biomechanical/article-abstract/137/2/024701/371210

This dataset contains the chat records from 15 implementations of *Nephrotex*. The chats have been labelled for the presence or absense of key concepts in engineering discourse that relate to particular *design moves* and *design justifications*. The data also includes marks for each student's final design report during the internship. The goal of this projects is to model performane on the report using data about what the students discussed throughout the internship.

Content

The virtual internship chat data is included in a csv file that contains the following:

1. "userIDs": a unique id for each student
2. "implementation": a unique id for each implementation
3. "Line_ID": a unique id for each chat utterance
4. "ChatGroup": a non-unique id for the team the chatter was in. String values are teams from the first half of the internship, numeric are teams from the second half. (Individuals switched teams halfway through).
5. "content": the content of each chat utterance.
6. "group_id": a non-unique numerical id for the team the chatter was in.
7. "RoleName": whether the chatter was a mentor or a player (i.e., student).
8. "roomName": a non-unique name for the internship activity the chatter was participating in.
9. "m_experimental_testing": Talk about using experimental techniques to understand the technical features of a design.
10. "m_making_design_choices": Talk about choosing a specification or characteristic for a design.
11. "m_asking_questions": Asking questions.
12. "j_customer_consultant_requests": Talk about justifying design choices by

stating that they should meet or exceed stakeholder (i.e., consultant) requests.
13. "j_performance_parameters_requirements": Talk about justifying design choices by referring to performance parameters or experimental results.
14. "j_communication": talk about justifying design choices by referring to facilitating communication among the engineers.
15. "OutcomeScore": a mark from 0 to 8 indicating the quality of the chatter's final design report. (0 is low quality, 8 is high)
16. "wordCount": the word count for the chat utterance

Project Objectives

1. Convert the data to team level statistics
2. Develop models to predict the final report marks at the team level using information about how the teams chatted. These models could include (but are not limited to): linear, ordinal, or logistic regression principal components analysis SVM Decision trees
3. Draw conclusions from the models as to how discourse features relate to report marks in this context. Consider the effect the mentor might have on these relationships.