**Submit**: project report (typed .. Word or PDF document)

**Collaboration**: yes – you can work alone, or in pairs

(**one report is acceptable for two students**)

Follow **the information discovery process** to investigate a dataset.

There are two possible datasets:

* horse-colic.arff Q: Which condition(s) leads to the death of the horse?
* thirdgrade.csv Q: What is related to good problem solving scores?

and two possible information discovery techniques (e.g., data mining, or visualization). So, there are four different possible projects: (1. horse-colic and datamining, 2. horsecolic and visualization, 3. thirdgrade and datamining, 4. thirdgrade and visualization).

**What to do (the project):**

* pick a dataset (just one)
* pick a technique (just one)
* follow the appropriate information discovery process to create a decision tree or visualization that ‘answers the question’
* **write a report** (similar to previous projects) that details the process that leads to the creation of the decision tree or visualization. Be sure to include the decision tree or visualization.

|  |  |  |
| --- | --- | --- |
| CS560-BJC | Information Discovery in EHR  Project #3: “Your Choice” | Spring-1 2020 |

Anytime you have questions about the data or the process, please post on the Discussion Board in Week 7.