



MINI STUDENT HACKATHON

The hackathon is limited to 2 hours.

A cleaned data set will be provided that has 10,000 rows and 21 columns. The first 20 columns are numeric input variables and the last column is the continuous dependent variable. The task is to build as good a predictive model as possible during the 2 hour period. The results will be measured by the r-squared error on a holdout data set, supplied when the team is ready. The team gets only 1 run on the holdout data, so use your best model!

The data is motivated by an particular data science problem that Monsanto had in modeling the relationship between many aspects of agriculture farm fields and the resulting yield of the crops. Aspects included things like soil conditions, fertilizers, herbicides, special seeds and weather. The data is synthesized but characteristic of this real problem.

Bring your laptops. Work in teams however you like.

Prizes are

First place team shares \$500

Second place team shares \$250

Third place team shares \$100