Overview of the D2K Learning Lab
Spring 2019

What is the D2K Learning Lab?
In the D2K Learning Lab, companies and community affiliates partner with Rice University to sponsor teams of students that work on a semester-long data science project.

- Student teams work with the sponsor’s data and on a sponsor-defined data science problem. D2K Lab faculty work with sponsors to help define projects.
- Teams of 4-6 students come from diverse disciplines and experience levels (e.g. undergraduate and graduate students).
- Teams are co-mentored by a sponsor mentor and a Rice data science faculty mentor.
- Teams produce a data science research or analysis report and deliver all software and scripts used in the report.
- Teams will exclusively use open-source software such as R and Python.
- Resulting Intellectual Property (IP) will be owned by the sponsor if a sponsored research agreement is signed.

What is the D2K Lab sponsored research agreement?
The sponsored research agreement is the mechanism under which affiliates can sponsor a D2K Learning Lab student team and own any resulting Intellectual Property.

- The Agreement is “turn-key”. Terms are not negotiable. During the pilot-stage of the D2K Learning Lab for the 2018-2019 academic year, we would appreciate feedback on this agreement and will work hard to meet the needs of our partners.
- For information on sponsorship fees for projects please contact us at d2k@rice.edu.

What types of projects are appropriate?

- Projects should be real-world, impactful projects that would be meaningful to the sponsor and motivating for the students.
- Projects should be challenging, but with well-defined objectives. It is recommended that the objectives be tiered in nature with some that are easily achievable within a semester and with others that are a stretch to motivate and challenge the students.
- While project objectives can focus on any aspect of the data analysis pipeline, the most effective projects will encompass aspects of each of the four stages of data science:
  1. **Data Wrangling** is the process of cleaning and unifying messy and complex data sets for easy access and data analysis.
  2. **Exploratory Data Analysis** (or Data Mining) visually explores data and finds interesting features or patterns in large data sets.
  3. **Data Science Modeling** uses statistical or machine learning models to address specific objectives.
  4. **Communication & Validation of Results** is critical for confirming data-driven discoveries and assessing the accuracy of data science models.
The best projects are organized in such a way that student teams can get started on day one.

- While student teams will need to learn discipline-specific terms and data complexities, the more these can be minimized or abstracted away, the more time student teams will be able to spend on the data science aspects.
- In cases where data-wrangling is a significant portion of the project, it is helpful to provide students with a subset of “cleaned” data that they can use as a model to begin thinking about how to address the project’s objectives.

What data is appropriate for the D2K Learning Lab?

- We encourage sponsors NOT to provide confidential data or divulge any confidential information to the student teams or faculty mentors.
  - Confidential data requires additional protections and training for students and faculty that may delay work on the project.
  - For 2018-2019 academic year projects, we will not be able to support projects utilizing confidential or protected data.
- A non-disclosure agreement for the students and faculty mentor is built into the sponsored research agreement. Rice will educate students and faculty as to the particular clauses of this agreement.
- We encourage sponsors to share data that they own but that is not proprietary or confidential. The data should not be personally identifiable and should not be data protected by HIPPA or FIRPA (such as health, medical, or educational data).

Timeline for Spring 2019 D2K Learning Lab Projects

Before semester starts:
- September to November – Discuss possible D2K Learning Lab projects with the faculty director (Genevera Allen - d2k@rice.edu).
- **November 1, 2018** – Project description submitted to the faculty director.
- **December 1, 2018** – Sponsored research agreement signed (firm deadline) and data for project delivered to the faculty director.

During Spring 2019 semester:
- **January 9, 2019 4pm - 5:30pm** – Orientation to the D2K Learning Lab and sponsors meet student teams.
- April TBD, 2019 – Spring D2K Learning Lab Showcase.
- May 1, 2019 – Final report and code due.

For more information: d2k@rice.edu