

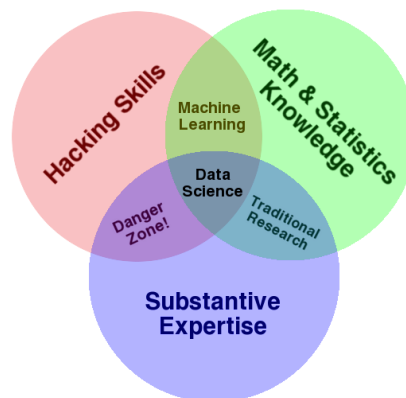
CSC-272
Data Mining
Fall 2023
Written Assignment 2

Feel free to use resources listed on [Datasets and Project Ideas](#) page of the course website.

You are strongly encouraged to discuss your project ideas and datasets with me during office hours. If you have a conflict, please [use this link](#) to schedule a meeting.

1. Identify group members for your course project. Working alone on the project is strongly discouraged. Group sizes should range from two to three students.
2. Drawing on datasets you and your group members identified in [Written Assignment 1](#) (Q3), reach a consensus on 1-3 datasets that you are *all* interested in working on together as a group.

Please be mindful of the scale of each dataset: it should be large enough to be appropriate for usage for a prediction task, but not so large that it is unwieldy (50 MB - 5 GB is generally recommended).



Data Science Venn Diagram

Reminder: The content of this course is going to cover ☒ Hacking Skills and ☒ Math & Statistics Knowledge but not ☒ Substantive Expertise (in a particular domain e.g. Business, Economics, Biology etc.). For that missing piece, I strongly encourage you to bring in knowledge from your GERs and other Non-CS/Non-IT interests into the course project.

3. Identify 3 interesting questions you and your group members want to answer using the dataset(s) you identified in the previous question. Use ideas listed on the [Datasets and Project Ideas](#) page of the course website for reference. At least one of the 3 questions should involve predicting a feature (column) from other features in the dataset.