

Software Requirements

Given the free-form nature of the machine learning capstone, the software and libraries you will need to successfully complete your work will vary depending on the chosen application area and problem definition. Because of this, it is imperative that all necessary software and libraries you consider using in your capstone project are accessible clearly documented. Please note that proprietary software, software that requires private licenses, or software behind a paywall or login account should be avoided.

Data Requirements

Every machine learning capstone project will most certainly require some form of dataset or input data structure (input text files, images, etc.). Similar to the software requirements above, the data you are considering must either be publicly accessible or provided by you during the submission process, and private or proprietary data should not be used without expressed permission. Please take into consideration the file size of your data — while there is no strict upper limit, input files that are excessively large may require reviewers longer than an acceptable amount of time to acquire all of your project files. This can take away from the reviewer's time that could be put towards evaluating your proposal. If the data you are considering fits the criteria of being too large, consider whether you could work with a subset of the data instead, or provide a representative sample of the data.

Ethics

Udacity's A/B Testing course, as part of the Data Analyst Nanodegree, has a segment that discusses **the sensitivity of data** and the expectation of privacy from those whose information has been collected. While most data you find available to the public will not have any ethical complications, it is extremely important that you are considering where the data you are using came from, and whether that data contains any sensitive information. For example, if you worked for a bank and wanted to use customers' bank statements as part of your project, this would most likely be an unethical choice of data and should be avoided.

If you have any questions regarding the nature of a dataset or software you intend to use for the capstone project, please send an email to **[**machine-support@udacity.com](mailto:machine-support@udacity.com)** with the subject "Capstone Project Dataset/Software Inquiry".