

Project Proposal

Due Monday by 23:59 **Points** 0

Your project will revolve around the question on **how a specific Data Science analysis method is challenged by Big Data properties such as volume, velocity or variety**. You are expected to find the specific method to investigate yourself. Sources for methods to investigate could be analyses you have performed previously, popular R or Python packages, tutorials, literature, problems you encountered when working in industry, etc. Make sure you pick a **very specific problem**, for example "normalizing date representations in varied data sources" or "parallelizing prior probability computation for text classification" and not "parallelizing scikit-learn" or "a solution to the halting problem".

The general challenges that arise from big data could be related to volume/velocity, e.g. the raw amount of data to be processed in a certain time requires parallelization of the approach over multiple computers. Challenges also arise from variety, where data can be expected to be dirty enough to create issues for analysis methods.



In this assignment, you will create a project proposal that describes the method and the big data aspect you would like to address.

Document Structure

- title
- an abstract, which you write last and only summarizes the document in a few sentences
- an introduction that motivates the topic and explains why it is interesting and relevant
- a related work section that points to the most relevant related work
- a section with a main scientific research questions
- a section with the approach, i.e. the main ideas that you want to try, **worked out to the point that these are understandable**.


See also this [guide on how to write technical papers](https://cs.stanford.edu/people/widom/paper-writing.html)  [_ \(https://cs.stanford.edu/people/widom/paper-writing.html\)](https://cs.stanford.edu/people/widom/paper-writing.html).

Formatting

You are required to use the ACM [two-column style](https://www.acm.org/publications/proceedings-template)  [_ \(https://www.acm.org/publications/proceedings-template\)](https://www.acm.org/publications/proceedings-template) for your paper. We recommend that you use Overleaf to write your paper in LaTeX. To simplify the process, we have created a [template on Overleaf](https://www.overleaf.com/read/tyzhgcprmrxd)  [_ \(https://www.overleaf.com/read/tyzhgcprmrxd\)](https://www.overleaf.com/read/tyzhgcprmrxd), that you can copy into a new project for yourself (Menu > Copy, available after creating an account and logging in).

Please limit your project proposal to 500 words.

Submission System

We use the EasyChair system to keep track of uploaded papers, reviews, meta-reviews, revisions etc. Submit your project proposal here: <https://easychair.org/my/conference?conf=bdds20>  [_ \(https://easychair.org/my/conference?conf=bdds20\)](https://easychair.org/my/conference?conf=bdds20) (You will have to create an account once). We only accept PDF documents. We will not accept proposals by email. Step-by-step instructions for EasyChair are available: [paper-submission-instructions.pdf](#) 