

Project

For this project you will work in groups of size 3, two of whom need to be CS majors and one majoring in the Life Sciences. Any deviation from this format needs to be cleared with me. Your project needs to be picked soon. You need to write a brief proposal. The proposal should include your name, e-mail addresses, date, title of project. It should provide a concise **problem statement, goals, motivation, hypotheses** (if any), **approach/methodology, data source(s), available software tools, and final deliverables**. The proposal is to be mailed to cap5510-f18@cs.fiu.edu or to cap5166-f18@cs.fiu.edu by Oct 8 before start of class.

Projects need to be research-oriented and/or result-oriented. I want to see some good results by the end of the semester. So start early. You are **advised** to email me an update of your progress on the 1st and 15th of every month until the end of semester and seek some feedback from me. Maintain a log file (or journal) with your activities on this project including: updates on your reading, progress on implementations and partial testing, ideas for future work, ideas that you may not be able to follow up, bug fixes, known current bugs in your code, organization of your program files and data files, etc.

At the end of the project, you will need to write a report (in docx or LaTeX format). It must include a short summary of your project. State clearly the following: your name, e-mail addresses, date, title of project, goals, hypotheses or assumptions, background with references and URLs, methods used (with references), what was implemented or achieved, summary of results, conclusions, possible future work.

Finally, prepare: (1) a 20-minute PowerPoint presentation of your work, (2) a short handout to distribute to your classmates, (3) web page describing your project, and (4) a zip-compressed file containing your (commented) source code, data, results, report, and webpage to be mailed to me. Your project should be completed and submitted by **November 19** before the Thanksgiving weekend because your presentations will start from Monday, November 26. Some sample projects are given below.

Metagenomics Pipeline with PluMA

1. We have a metagenomics pipeline called PluMA that takes NGS data on samples and converts them to abundance matrices and does much more. Your goal is to take more tools such as QiiME and to add plugins to the software pipeline and then show that it can be used for performing metagenomics analysis. More information on PluMA can be found here: <http://biorg.cis.fiu.edu/pluma/>

Deep Learning for Network Biology

2. There has been considerable interest recently in using Deep Learning for Network Biology. Your project would be to replicate some of the techniques from this domain with different data sets. A possible question is to find ways to apply it to Metagenomics. More details can be found at <http://snap.stanford.edu/deepnetbio-ismb/>

Mining for Relationships from Microbial Ecology data

3. Microbial organisms produce metabolites, which are often a basis for interactions between microbes or between microbes and the host they infect. A recent paper has found ways to infer microbial relationships by doing simulations. A possible project is to figure out how to use this data to validate predictions of microbial interactions from metagenomic data. The paper mentioned above is at: <https://www.nature.com/articles/ncomms1597>