CSCI-B565: Data Mining

Homework Assignment # 3

Assigned: 02/24/2018 *Due: 03/03/2018, 11:59pm, through Oncourse*

One project proposal, 50 points in total. Good luck!
Prof. Predrag Radivojac, Indiana University, Bloomington

Mini project proposal. (50 points) Form a group of up to three individuals and submit a proposal for the mini project. Groups with less than three members are encouraged only if you are conducting research with a particular Indiana University faculty. Each person in a team must submit an identical copy of the group project.

Instructions for proposal writing

The project needs to be interesting and/or important, but also doable in 6-8 weeks by at most three people. You are not only allowed but also encouraged to combine this project with your research tasks with different faculty at IU. However, it needs to be made clear to the instructor what is the part of your work that has been done specifically for the purpose of the class project, even if your results later become useful for another project.

There are many ways to carry out an interesting project. The only bottom line is that the project needs to non-trivially involve concepts related to CSCI-B565. For example, you may desire to repeat other people's published work to verify whether the results stand, you may want to explore various concepts from class and do detailed analysis, you may want to build a tool that works (e.g. a mini recommendation system), you may want to compete in Kaggle and report how well you did (in this case, you cannot simply download data, import it to a tool and make predictions; that would fall under the domain of trivial). The best approach is to do what you are motivated to do.

You can use various software libraries. The quality of the project will be judged based on the quality of your ideas, interestingness of the project, and the work you carried out.

The format of the project proposal is listed below.

1. Project title.
2. Names of the team members and their Indiana University emails.
   (a) If the project is under the guidance of a particular Indiana University faculty member, list the professor (or professors) working with you.
3. Objectives and significance (1-3 paragraphs)
   (a) Describe what the goal of the project is, why is it important, and your motivation for doing it.
4. Background (1-2 pages)
   (a) Introduce all important concepts and background information.

   (b) Search the literature and describe previous work on this problem.

   (c) If there exists previous work on the problem, describe what makes your work particularly interesting.

5. Proposed approach (2-4 pages)

   (a) Describe your data and how you will obtain it.

   (b) Describe your proposed method and implementation. Be as detailed as necessary.

   (c) Describe evaluation strategy. Be as detailed as necessary.

   (d) Describe expected outcomes with a fall back on option in case the initial idea fails.

6. Individual tasks (1-3 paragraphs)

   (a) Describe what each member of the team will do and justify why each person is necessary. In the project report (at the end of semester) you will have to be specific about the roles of everyone.

   (b) If the project is a part or extension of some other project you have carried out (or will), explain all relevant details and what is the work that differs from what you already have done or are doing for another purpose.

7. References

   (a) List books, scientific papers, web sites etc. that you referenced in the proposal body.

Technical details:

1. Use 1 inch margins from each side.
2. Use 11pt or 12pt font size.
3. Use standard font types such as Times New Roman, Arial, or Latex default fonts.
4. Use single or 1.5 line spacing.

Homework Directions and Policies

Submit a single package containing all answers, results and code. You submission package should be compressed and named firstnamelastname.zip (e.g., predragradivojac.zip). In your package there should be a single pdf file named main.pdf that will contain answers to all questions, all figures, and all relevant results. Your solutions and answers must be typed[1] and make sure that you type your name and IU username (email) at the beginning of the file. The rest of the package should contain all code that you used. The code should be properly organized in folders and subfolders, one for each question or problem. All code, if applicable, should be turned in when you submit your assignment as it may be necessary to demo your programs to the associate instructors. Use Matlab, Python, R, or C/C++.

Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rules:

on time: your score $\times$ 1

1 day late: your score $\times$ 0.9

2 days late: your score $\times$ 0.7

3 days late: your score $\times$ 0.5

4 days late: your score $\times$ 0.3

5 days late: your score $\times$ 0.1

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points.

All assignments are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged; e.g., web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously! For detailed information see Indiana University Code of Student Rights, Responsibilities, and Conduct.

[1] We recommend Latex; in particular, TexShop-MacTeX combination for a Mac and TeXnicCenter-MiKTex combination on Windows. An easy way to start with Latex is to use the freely available Lyx. You can also use Microsoft Word or other programs that can display formulas professionally.