NYU | STERN

DATA SCIENCE FOR BIZ ANALYTICS
Spring 2019

Instructions for term project

Prof. Provost & Carlos Fernández

For your term project: (see the syllabus for the due dates for the deliverables)

You will mine actual data for a problem of interest. These could be data from a problem from your current job, something of interest to the school, data acquired from the web, etc. You will design the data mining task, mine the data, and describe your results. You also will research existing solutions to the problem, if any have been proposed or documented. Your own data and results need not be on par with actual industry results; the goal is for you to get as realistic a hands-on experience as possible, given the constraints of what you have learned.

In writing up/presenting your research, think of yourselves as analysts employed by or retained by a company (large or small) or by a funding source (e.g., a VC firm or incubator), who wants to understand the state of the art for using data mining for the task in question. Review what has been done to date on your problem. Consider as an example predictive analytics for on-line advertising: A VC firm considering funding on-line ad networks or ad-tech startups would need to understand the state of the art in using data mining for targeting on-line advertising, when considering an idea for applying data mining. Don't worry too much about coming up with a novel idea. It is more important to develop the idea well (within the scope of what we've discussed in class).

You should use the "data mining process" to structure your research and writeup. Keep in mind that it may be ineffective simply to proceed linearly through the steps, and this may need to be reflected in your analysis. You should interact with me and the course assistants from the preparation of your initial ideas through your write-up, as a consulting group would interact with a firm or funding source in preparing a research report. Use your imagination, prior experience, or ask us to help to fill in any gaps between the material available and what you would be able to find out if you actually could interact with the client firm.

Deliverable #1: By **Feb 27th** you will submit your choices for teams and initial ideas for projects. Teams will comprise 3-4 students; in class we will give you instructions on how to form your teams. Initial ideas can simply be a paragraph about what you are thinking you might do.

Deliverable #2: No later than **March 14th** you will submit a **proposal** for your project. This should give as much detail as possible your ideas, so that we can give you feedback. Include in your proposal your ideas about: What is the exact (business) problem? What is the use scenario? What precisely is the data mining problem? Is it supervised or unsupervised? What is a data instance? What might be the target variable (if supervised)? What features would be useful? How exactly would the "technical" solution solve the business problem? Etc.

Deliverable #3: By **April 5th** you will submit a status report ("Project Update"), including preliminary results and any issues that you are facing in developing your project.

Deliverable #4: No later than **TBD** you will submit your final write-up, which should include the information detailed below, in approximately the order given. Your write-up need not have corresponding sections or bullet points, but I should be able to find the information without searching too hard. Be as precise/specific as you can. The write-up should be about 20 1.5-spaced pages, plus any appendices you would like to include. Use external sources where appropriate, and provide clear citations and bibliography. All group members should contribute to the analysis and write-up. The report should include an appendix describing the contributions of each team member.

You will get the most out of the project if you interact with us during the development of your ideas. Talk to us especially before choosing one of the business problems we cover in class. And please feel free to come talk to us about your ideas as often as you'd like. Please do not choose stock/index prediction or market forecasting (talk to Foster).

(see other side)

Your project write-up should include the information detailed below, in approximately the order given. Your write-up need not have corresponding sections or bullet points, but we should be able to find the information without searching too hard. Be as precise/specific as you can.

Business Understanding (take this seriously)

- Identify, define, and motivate the business problem that you are addressing.
- How (precisely) will a data mining solution address the business problem?

(NB: I'd like to see a good definition/motivation of the business problem and a precise statement of how a data mining solution will address the problem. It's not so important that the hands-on results match perfectly. It's more important that you have the experience of working through a realistic problem definition.)

Data Understanding

.

- Identify and describe the data (and data sources) that will support data mining to address the business problem. Include those aspects of the data that we routinely talk about in class and/or in the homeworks.

Data Preparation

- Specify how these data are integrated to produce the format required for data mining.

(NB: data preparation can be time consuming! Get started early. Talk to the CAs or Prof if you need advice.)

Modeling

- Specify the type of model(s) built and/or patterns mined.
- Discuss choices for data mining algorithm: what are alternatives, and what are the pros and cons?
- Discuss why and how this model should "solve" the business problem (i.e., improve along some dimension of interest to the firm).

Evaluation

- Discuss how the result of the data mining is/should be evaluated. How should a business case be developed to project expected improvement? ROI? If this is impossible/very difficult, explain why and identify any viable alternatives.

Deployment

- Discuss how the result of the data mining will be deployed.
- Discuss any issues the firm should be aware of regarding deployment.
- Are there important ethical considerations?
- Identify the risks associated with your proposed plan and how you would mitigate them.
