Term Project

As part of this course you are required to complete a term project. The project has two components (i) a presentation worth 10% of the grade in session 10 and (ii) a project report worth 10%.

The project for the class involves finding a business or government dataset and applying the analytical methods described in the course to the data to derive useful insight for decision-making. You should do this project in teams of 5 students.

Proposal: Submit 1 page document listing names of team members and idea for project
Week 10: Submit presentation slides (PowerPoint or PDF file) to Canvas
Week 10: Submit project report to Canvas

Ideas for Project

You can use private corporate data or publicly-available data from the Internet. If you use corporate data, the data can remain confidential but you will have to present your results to the class.
The data analysis could involve for instance finding out what factors predict borrowers' defaulting on their loans, how to predict which stations in a manufacturing line are responsible for product defects and predicting which customers will respond to an offer. Popular ideas in the past have used stock-market prices, data from sport leagues (NFL, NBA, etc) and data from Internet databases. These include examples such as how online retailers price differentiate (from sites such as bizrate.com), how airlines make pricing decisions (makemytrip.com), how consumers bid in a given auction or a series of auctions (baazee.com), the degree of interest a news piece generates (blog ranking sites), products reviews on Amazon, keyword search trends available from Google (trends.google.com) and data-sets from UCI AI lab website.

The deliverables: a project report and a 10 minute presentation to the class.

Guidelines for the presentation and report are provided below.

Guidelines for Term Project Presentations

Making an effective presentation is an important skill. Here is my view of the **perfect presentation**:

1) Presentation material has been uploaded to dropbox in advance

2) Presenter is energized and sounds enthusiastic – not droning

3) The presentation is useful, focused and delivers new insights for the audience

4) The slides contains short cues for the speaker – not full sentences

5) Slides contain graphical exhibits

6) Humor is used to lighten the mood occasionally

7) Team members maintain eye contact with the audience, glancing briefly at the slides but not reading from them

8) Team is smartly dressed (business casual)

9) Presenter is very familiar and comfortable with the content of the presentation

10) Presenter is focused on communicating facts or concepts in a crisp style without undue repetition

11) Presentation ends in 10 minutes.

Term Project Report
1. Executive summary
2. Explain the business idea, why it's important, and the data source
3. Data summary, description, visualization.
4. The type of analysis you conducted and why. I suggest trying 2 different techniques.
5. (a) Show results from 2 benchmarks:
   (i) benchmark accuracy without pre-processing data.
   (ii) report the proportions of the class variable.
   (b) Indicate which pre-processing steps you undertook, show any change in model accuracy due to pre-processing.
   (c) Iterate through step (b), trying different pre-processing steps, selecting different columns, binning, resampling etc. Report results – can you beat both benchmarks? Indicate if specific pre-processing steps were particularly valuable in improving the accuracy of the data mining methods.
6. (a) What did you learn about data-mining from your analysis – list some takeaways (b) Interpret and analyze your results to help business managers understand the implications and actions that follow from the analysis. (c) Other recommendations.

The write-up should be about 10 pages (not including all relevant printouts, tables, references and graphs) but the length is just a guideline.