Group Project

As part of this course, you are required to complete a group project. The project has two components (i) a presentation in week 14 or 15 worth 10% of the total course grade and (ii) a project report worth 15%.

The project 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 or 6 students.

Deliverables:

- Proposal: Submit a 1-page document listing names of team members and your idea for project.
- Week 14: Submit presentation slides (PowerPoint or PDF file) and give a 20-minute presentation in class.
- Week 10: Submit a 12-page (or less) project report.

Ideas for Project

In the one-page proposal, you should briefly specify your target problem, motivation, data source with a simple data description, and what you plan to do in this 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 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 datasets from UCI AI lab website.

Guidelines for the presentation and report are provided below.

Guidelines for Group Project Presentations

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

1) Presentation materials have been uploaded to eLearning in advance.

2) The presenter is energized and sounds enthusiastic – not droning.

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

4) The slides contain 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) The presenter is very familiar and comfortable with the content of the presentation.

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

11) The presentation ends in 20 minutes.

Guidelines for Group Project Report

1. Executive summary

2. Introduction: Explain the business idea, why it's important, and the data source

3. Data: Data summary, description, visualization.

4. Analysis: The type of analysis you conducted and why. I suggest trying 2 different techniques.

5. Results:

    (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. Discussion:

    (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 12 pages (not including references), 1.5 line spacing, font 11.