

ISM 4545

Project Guidelines

Summary

- This project focuses on data-driven decision making. Your goal is to provide business intelligence analysis for the managerial decisions of an organization.
- To accomplish this, you will need to identify a dataset, thoroughly analyze it using the techniques and processes covered in class and provide an executive level report as your final deliverable.

Summary (cont.)

- Teams of 1 – 4
- At least one dataset coming from an external data source
 - Examples are on the next slide
- 3 high level business questions to answer from the data
 - Must be answered using the process and the techniques that we've covered in class
 - Detailed steps on slide #5
- Deliverable is a technical report with an executive summary of your findings and suggestions
 - Detailed grading criteria on slide #7

Where can I get data from?

- There are numerous publicly accessible (government, industry, academia, sports, etc.) data repositories on the web that are suitable for this project
- Some examples:
 - Competitions:
 - Kaggle
 - <http://connect.informs.org/oratc/2017problem>
 - Kdnuggets
 - U.S. Bureau of Labor Statistics
 - UC Irvine Machine Learning Repository
 - Research centers (Microsoft, IBM, etc.)
 - <https://cds.cdm.depaul.edu/resources/datasets/>
- Try googling “data mining data sets”

Steps

1. Create your team
2. Find the data
3. Identify at least 3 business questions that can be answered from the data
4. Conduct data analytics and mining to answer these questions
 - Detailed info on next slide
5. Create and submit your project report

Bottom up approach

Process / Tasks

- For reporting purposes, you can write-up your analysis based on the CRISP-DM approach, which includes the steps: Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation.
- Possible specific tasks involved in the analysis are:
 - Pre-processing the data (cleaning, transforming, etc.)
 - Descriptive statistics and exploratory analysis
 - Cluster analysis, if appropriate
 - Classification and prediction, if appropriate
 - Association rule mining, if appropriate
 - Social network analysis, if appropriate
 - Text mining, if appropriate
 - Optimization, if appropriate

Grading Criteria

- You will be graded on your ability to demonstrate your grasp of data analytics concepts and how they can be applied in a business context. Grades will be determined based on your final project report. The report must not exceed 15 pages, including all the images, tables and references. The report should include (but not limited to):
 - A section that clearly describes and discusses the business context and the series of business questions that you have selected to answer
 - A section that provides information about the data and the data collection (where did you get the data, how many records / attributes, what kind of dataset, etc.)
 - A summary of the data preparation and pre-processing steps taken
 - Descriptive statistics and exploratory analysis (with Tables and Figures, if appropriate)
 - For each business question, a discussion on the model and the methods selected, with a visual representation for each model (i.e. screenshots)
 - A results and conclusion section with interpretation of the results within the context of your business questions.
 - An executive summary of your findings and suggestions.