

Course Project Guidelines

General Description

This project requires students to apply data-mining techniques learned in class to a business problem using real-world data. The data can be collected from (i) an organization's database (ii) a publicly available source

Guidelines for Project Report

The final report should include the following items:

1. Background information about the data, including related organization/industry information.
2. The objectives of the project and the solution techniques employed (e.g., decision trees, clustering, etc.). Explain current/none-data-mining methods issues, why your solution is better than current/none-data-mining operational processes and why you choose these techniques over the others. So comparing with existing/none-data-mining methods, you can argue/claim that 1) it save cost or 2) it increase sales or 3) it increase efficiency/ make better decisions. Of course, money is one of the metrics to measure it. Other arguments/evidences/metrics for business improvement are OK as well.
3. Data description: Summarize the data in a three-column table. List the name of the attributes in the first column, the data type (numeric, categorical, etc.) in the second column, and the descriptions of the attributes in the third column. Describe each attribute in a way such that students in this class can understand. Indicate the number of total records. If the data is proprietary, mask the confidential elements of the data. If there are missing values, explain how you handle them. Data mining process: Including data cleansing, attribute selection and transformation, training and testing process. Interpretation of the results. For example, show your final decision tree and rules generated from the tree.
4. Managerial and other practical implications and insights of your project. If applicable, recommendations to the organization based on your study.

You are required to submit the data for the project as supplemental materials, including the data source. If the data have been processed for the project, both the original and the final versions of the data should be submitted.

Format and Schedule

- Submit a one-page plan that includes the following items on or before week 3
 - The objectives of the project.
 - Data-mining techniques to be used.
 - A brief description of the data, including the data source, a list of attributes and the number of records.
- The final report (in Microsoft Word), should be between 8 and 12 pages, 1.5 line spacing, all things included.