

Assignment 6 (15%): Real-World Data Mining Competition. (Due on 16:59:59, Dec. 21, 2016)

Since we have learned a variety of data mining techniques so far, why not try to participate in a real-world data mining competition now in order to test your ability and see how data mining techniques are applied in the real-world.

In this assignment, you are asked to participate in an ongoing real-world data mining competition held by [kaggle](#) (a place where you can compete with world-class data miners on real data mining applications, and at the same time possibly earn a handsome prize). This year, you are asked to design and implement data mining methods to handle the problem: [Santander Product Recommendation \(1st place - \\$60,000\)](#). Please participate in the competition alone and teamwork is NOT allowed.

You can click the [Santander Product Recommendation](#) to get all the information you need to accomplish the competition, including how to register, how to get the data, how to evaluate your method. To be explicit, we give a brief guidance for the newcomers of kaggle, which is described below.

Task

1. Modify the Kaggle Team Name (the name which will be shown in the [leaderboard](#)) in [My Team](#) web page as your student number, e.g., MG1633001. You can use any account name as you like, but the Team Name that will appear in the leaderboard must be your student ID. If your Team Name is not your student ID, your rank will NOT be recorded by us at deadline, and thus your score for rank will be zero.
2. Participate in the competition and get higher rank as possible as you can.
3. Write a report to describe your methods and implementation.
4. Submit your report, code, ReadMe, as well as the prediction file to FTP.
5. You can use any popular machine learning/data mining toolboxes in the competition. Some good ideas can be found in the [forums](#).

What should you submit to FTP

Pack all the files needed to be submitted, i.e., report.docx, code.zip, ReadMe.txt, prediction_on_test.csv. Before submitting your assignment to FTP, please read Submission Requirement and Description section above carefully and obey it.

1. The prediction_on_test.csv should be in the same format as your have submitted to kaggle. The report should contain your final rank and result before deadline (2016-12-21 16:59:59) in the leaderboard, your name, student ID, and e-mail address. Name this pack using your student ID, e.g., MG1533001.zip.