Final Project – Proposal

- **Identify machine learning problem that you are interested. You have the freedom to choose any problem (even the one proposed by yourself). But if the problem and the solution/code are very similar to the assignments, you will receive less credit due to difficulty level.**

The problem should be solvable using methodologies covered in this class. You could either find a problem by yourself with the instructor's permission or identify a topic below.

1. predict whether a college is private or public (clustering problem)
2. predict how much a customer would spend with the company (regression)
3. predict whether a customer would end his/her business with the bank (classification)
4. predict whether a person would click on a Facebook AD (classification)
5. predict whether a customer would end his/her business with the bank (classification, same dataset with task 3)
6. predict student math subject performance (regression)
7. MNIST handwritten digits recognition (multi-class classification)
8. Fake news recognition (classification)
9. Kernel methods for any above regression/classification problem

- **Two students in a team. Single member teams (not encouraged!) should commnunicate with instructors to explain.**
- **No restriction to use machine learning libraries. The grade depends on the implementation quality. But without using off-the-shelf tools usually will receive higher credit.**
- **Combination of different algorithms will receive higher credit, e.g., a pipeline with PCA or other feature pre-processing techniques combined with regression/classification.**
- **Each team must give the contribution of each team member.**
- **Each team will have 8 minutes to present, 2-3 minutes for Q&A.**
- **You need to submit your code and slides in the blackboard.**

The slides should include three components:

1. **Problem description**. It may be consisted of background, the importance, benefits of solving the problem, challenges, etc.
2. **Data understanding.** Understand the data and features by visualization or any quantitative analysis tools.
3. **Approach.** The machine learning technologies you are going to use to solve this problem. It is also to identify some potential challenges you may confront in the project and describe how you plan to address it. Teams using machine learning tools like Scikit-learn should explain more details about your understanding of the tool, what each parameter means, how you tune them, etc.
4. **Results.** Report the results with your conclusive remarks.