

# ASSIGNMENT 1

Weight: 40%

Submissions: A report (PDF) and a zip file (including code and datasets) on Brightspace.

The purpose of this assignment is to practice how to use data mining and machine learning to solve real-world problems. You will need to identify the target problem yourself. You can choose any project (**Note: Not related to real estate and stock market**) as long as it is legal. This assignment is a group project, and each group should have four members. Each group only needs to submit one solution.

Your pdf report should clearly detail how you carried out the experiment to address your targeted problem and show the results you got.

1. Your report should be written in Overleaf or Word, and use the provided template: <https://www.acm.org/publications/proceedings-template>.
2. It should be a human-readable document (e.g. do not include code)
3. The final report is expected to be 4-6 pages including references (4 pages are preferred).
4. You should provide student numbers instead of address in the provided template.
5. Use clear headings for each section.
6. Include tables and figures if needed appropriately, such as giving captions, describing your figures or analysing the results provided in your tables in your text etc.
7. The final report filename should be "Comp3027J\_GroupXX" (e.g. Comp3027J\_Group01)

In your report, it is recommended to discuss the following essential topics, but not limited to these topics:

1. What is the real-world problem addressed and why it is important.
2. Dataset selection (collection) and Data pre-processing.  
Where you find your data (or how do you collect the data and create your dataset)?  
How do you analyze your data?  
how to pre-process your data to fit your solution?  
Any challenges with your dataset?  
etc.
3. Methodology  
Any machine learning algorithm can be used (not limited to the algorithm we have learned).  
Creativity is encouraged.  
Be careful, a sophisticated approach with little description and explanation will receive little credit.
4. Evaluation  
Elaborate your experiment, such as splitting dataset, K-fold;  
Compare your solution with benchmarks in literature;  
Evaluation metrics for your task;  
Analysing your results etc.

In your zip file, you should include your code and dataset. And please make sure to clean up your code to make the results reproducible. If its size exceeds the Brightspace limit, it needs to be submitted via a USB key.