Take-home Examination: Part I-Stock Clustering

Due Date: **17 December 2021**

Marks: 15%

Objectives

- Demonstrate your problem formulation skill
- Develop a conceptually sound data mining solution using tools, e.g. Weka 3, or SciKit , or by programming using a familiar programming language

---

Brief Description

By using the provided HK Stock data consisting of six stocks (code 1, 11, 293, 857, 13 and 23), formulate and implement a stock clustering solution, i.e., to do a kind of clustering of the provided stock data, and report the clustering results. Note that you are expected to conceive a reasonable clustering task on your own! Note also that there exists missing data from stock 857 as its IPO (initial public offering) date is later at 7 April 2000. The marking criteria are as follows.

Formulation and Solution	70%
Analysis and Presentation	30%

This small project is the first part of the take-home examination. Only a simple and conceptual sound solution based on traditional clustering methods is expected but your problem formulation has to be thoughtful and with practical usage. Note that you are supposed to work on the provided six stocks' data only and are NOT expected to use other stock data to develop your solution. The remaining part of the take-home examination will be written questions (no programming) to be released later, potentially following the context of stock data mining and

data warehousing and extending to other problems and/or issues of DM/DW.

Tips:

1. Intra-stock vs inter-stock data mining
2. Conceptually sound solution is good enough. Hence, you are recommended to simply use the daily closing price of the six stocks to develop the solution and consider the use of other information like daily high/low and transaction volume to possibly extend your solution in the future. Similarly, the use of sophisticated clustering methods can be considered for future work.

What to hand in

- A short (<=6 pages) written report summarizing the work (formulation and solution) you have done, the analysis you have made, and the clustering results you have obtained.
- The source codes (if any) or design (if using tools) of the system.

References

1. Tesla Stock Price Data@Kaggle