Queen Mary University of London

Assessed Project

Submit online via QMplus before 23:59 London Time on Monday 11th September 2023

ECOM193 Statistical Machine Learning in Finance

This project carries a maximum of 100 marks. A 50% weighting will be applied to your overall ECOM193 module score for completion.

You are required to submit a typed document in Word or PDF format containing written analysis together with any supporting tables, graphics and code you think is necessary.

The details of what is required for this project are given on the next page.

This project has an maximum word count of 2500 words **excluding** computer output and any R code that you choose to submit.

This project **must** be your own work presented in your own words.

Examiner: Dr R.A. Saldanha

Similarity of US Technology Stocks

Cluster analysis is a technique used in unsupervised machine learning to group a set of objects in such a way that objects in the same group or *cluster* are more similar to each other than to those in other clusters.

1. Describe suitable methods for undertaking hierarchical and non-hierarchical cluster analysis. In particular, explain how different distance measures and/or clustering algorithms might alter the results and how choice of the number of clusters might be made.

   [35 marks]

2. The datafile `ustech.csv` consists of daily closing prices (USD) from January 2019 to June 2021 on 21 technology companies appearing in the Nasdaq 100 index.

   i. Explore the data, individual securities and their characteristics.

   ii. Use cluster analysis to determine if meaningful groups can be formed. Pay particular attention to data scaling, i.e. consider carefully if some form of data standardization is necessary.

   iii. What conclusions (if any) do you draw from your analyses?

   [45 marks]

3. What are the main disadvantages of cluster analysis? What do you deduce from your above analyses in light of these considerations?

   [20 marks]