# Dublin Business School

# Assessment Brief

# Assessment Details

|  |  |
| --- | --- |
| Module Title: | Machine learning |
| Module Code: | B9DA101 |
| Module Leader: | Dr Shahram Azizi |
| Stage (if relevant): |  |
| Assessment Title: | CA two |
| Assessment Number (if relevant): |  |
| Assessment Type: |  |
| Restrictions on Time/Length : | Submission before deadline |
| Individual/Group: |  |
| Assessment Weighting: |  |
| Issue Date: |  |
| Hand In Date: |  |
| Planned Feedback Date: |  |
| Mode of Submission: | Online |

**Guideline:**

* This CA assesses students on time series analytics, unsupervised learning and multivariate analytics in machine learning.
* All questions are mandatory.
* Use R/Rstudio to solve questions and perform analytics.
* Any submission after deadline will not be considered and scored.

**Question1** (each part 10 marks)

Use dataset available on <http://www.stat.ufl.edu/~winner/data/clotthes_expend.csv> , apply time series analysis, consider **sales.b** as your time series variable:

(a) Validate the assumptions.

(b) Fit the optimized model for **sales.b** and provide the coefficient estimates for the fitted model.

(c) What is the estimated order for AR and MA?

(d) Forecast h=10 step ahead prediction of the original time series.

**Question 2** (each part 20 marks)

Use the dataset available on

<http://www.stat.ufl.edu/~winner/data/iran_rock.csv>,

1. Perform ANOVA and summarize the output.
2. Load the dataset available on <http://www.stat.ufl.edu/~winner/data/esp_studies1.csv>,

Apply PCA, and identify the important principle components involving at least 80% of dataset variation. Explain your decision strategy.

Note: Technical support is available to student between **0930- 1700 hrs only**. There is no technical support after 1700 hrs. It is your responsibility to ensure that you allow time to troubleshoot any technical difficulties by uploading early on the due date.