# 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 one |
| Assessment Number (if relevant): |  |
| Assessment Type: |  |
| Restrictions on Time/Length : | Submission before deadline |
| Individual/Group: |  |
| Assessment Weighting: |  |
| Issue Date: |  |
| Hand In Date: | 30th July 2018 |
| Planned Feedback Date: |  |
| Mode of Submission: | Online |

**Guideline:**

* This CA assesses students on advanced data analytics techniques in machine learning namely an overview of main objectives and definition in machine learning, the most popular classification algorithms, time series forecasting.
* All questions are mandatory.
* Use R/Rstudio to solve questions and perform analytics.
* Any submission after deadline will not be considered and scored.

**Question 1: (each part 5 marks)**

1. What is the predictive analytics?
2. What is the main purpose of implementation of machine learning algorithms?
3. List the main criteria to select an appropriate machine learning algorithm before processing data.
4. Express the main differences between supervised versus unsupervised learning algorithms.
5. What are the main steps to design an optimal logistic classifier?
6. What are the main steps to design an optimal SVM classifier?

### Question 2: (each part five marks)

Using the dataset available on, <http://data.princeton.edu/wws509/datasets/cuse.dat>,

Propose the best classifier using Bayes Naïve and logistic regression. To do so, consider ‘wantsMore’ as the output variable:

1. Train the model using 70% of this dataset.
2. Predict the test set using the trained model.
3. Provide the confusion matrix and obtain the probability of correctness of predictions.
4. Write the functional form of your proposed classifier.

**Question 3: (each part five marks)**

Perform classification task using a linear SVM classifier to correctly classify ‘wool’ attribute in **warpbreaks** dataset. To do so,

1. Train your model using 80% percent of dataset and test the model using the remaining 20%.
2. Compute the confusion matrix and evaluate the performance of classifier.
3. Change the kernel of SVM to ‘radial’, and compute confusion matrix
4. Which kernel is preferable?

### Question 4:

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

1. Validate the assumptions using graphical visualization.

**(10 Marks)**

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

**(5 Marks)**

1. What is the estimated order for AR and MA?

**(5 Marks)**

1. Forecast h=10 step ahead prediction of **sales.b** on the plot of the original time series.

**(10 Marks)**

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.