

Module 36 Machine Learning

Module title	Machine Learning
Module NFQ level (only if an NFQ level can be demonstrated)	8
Module number/reference	BSCH-ML
Parent programme(s)	Bachelor of Science (Honours) in Computing Science
Stage of parent programme	Award stage
Semester (semester1/semester2 if applicable)	Semester 2
Module credit units (FET/HET/ECTS)	ECTS
Module credit number of units	5
List the teaching and learning modes	Direct, Blended
Entry requirements (statement of knowledge, skill and competence)	Learners must have achieved programme entry requirements.
Pre-requisite module titles	BSCH-FC, BSCH-LA, BSCH-NO, BSCH-PAS
Co-requisite module titles	None
Is this a capstone module? (Yes or No)	No
Specification of the qualifications (academic, pedagogical and professional/occupational) and experience required of staff (staff includes workplace personnel who are responsible for learners such as apprentices, trainees and learners in clinical placements)	Qualified to as least a Bachelor of Science (Honours) level in Computer Science or equivalent and with a Certificate in Training and Education (30 ECTS at level 9 on the NFQ) or equivalent.
Maximum number of learners per centre (or instance of the module)	60
Duration of the module	One Academic Semester, 12 weeks teaching
Average (over the duration of the module) of the contact hours per week	3
Module-specific physical resources and support required per centre (or instance of the module)	One class room with capacity for 60 learners along with one computer lab with capacity for 25 learners for each group of 25 learners

Analysis of required learning effort		
	Minimum ratio teacher / learner	Hours
Effort while in contact with staff		
Classroom and demonstrations	1:60	18
Monitoring and small-group teaching	1:25	18
Other (specify)		
Independent Learning		
Directed e-learning		
Independent Learning		57
Other hours (worksheets and assignments)		32
Work-based learning – learning effort		
Total Effort		125

Allocation of marks (within the module)					
	Continuous assessment	Supervised project	Proctored practical examination	Proctored written examination	Total
Percentage contribution	70%			30%	100%

Module aims and objectives

This module aims to empower learners to perform a wide range of machine learning tasks including but not limited to classification, prediction, regression, clustering, and association rule learning. Learners develop a deep understanding of these techniques and be able to apply them to a range of data to produce meaningful results.

This aim will be met through the pursuit of the following objectives

- To familiarize learners with a number of machine learning tasks including but not limited to classification, prediction, regression, clustering, and association rule learning.
- To equip learners with skills to represent these tasks with appropriate notation to express their familiarisation.
- To assist learners in expanding their programming competencies to include a programming language and tool suitable for performing machine learning tasks.
- To support learners in the sequence of: identifying the underlying data, classifying the problem type, selecting the appropriate machine learning algorithm, selecting the appropriate tool of implementing a customized tool to perform the machine learning task.
- To expose learners to a dataset concerning real world problems.
- To provide learners with a framework to evaluate the performance of the techniques on the tasks.

Minimum intended module learning outcomes

On successful completion of this module, the learner will be able to:

1. Apply a range of appropriate machine learning techniques to problems
2. Critically evaluate the different models of machine learning in respect of their use, accuracy and performance
3. Implement various data machine learning algorithms and techniques
4. Employ appropriate machine learning tools on large datasets to generate information
5. Analyse machine learning results and determine where improvements can be made.

Rationale for inclusion of the module in the programme and its contribution to the overall MIPLOs

As the volume and variety of data grows in society learning effectively from this data becomes an increasingly complex problem. Statistical analysis offers us limited ability to express the underlying patterns that exist in data. Machine learning allows us to discover, express and encode more complex decision boundaries than is possible through statistics.

Data analysis is fast becoming one of the key growth areas in computer science and machine learning algorithms provide excellent tools in performing data analysis tasks. Offering this module as part of this programme grants learners access to a large growth area in Computer Science.

Appendix 1 of the programme document maps MIPLOs to the modules through which they are delivered.

Information provided to learners about the module

Learners receive a programme handbook to include module descriptor, module learning outcomes (MIMLO), class plan, assignment briefs, assessment strategy, and reading materials.

Module content, organisation and structure

AI Introduction

- What is Machine Learning
- The Foundations and History of Machine Learning
- State of the Art

Data and Classification of Machine Learning Algorithms

- Types of Data
- Classification of Machine Learning Algorithms

Ensemble

- Random Forest
- AdaBoosting

Regression

- Linear Regression
- Logistic Regression

Bayesian

- Naive Bayes
- Bayesian Network

Decision Trees

- C4.5

Dimensionality Reduction

- Principle Component Analysis

Instance Based

- K-Nearest Neighbour
- Self-Organizing Map

Clustering

- K-means
- K-medoids
- DBSCAN

Deep Learning

- Unsupervised Pertained Networks
- Convolutional Neural Networks

- Recurrent Neural Networks
- Recursive Neural Networks

Module teaching and learning (including formative assessment) strategy

The module is taught as a combination of lectures and lab sessions. The lecture sessions discuss and assist learners in exploring the theoretical underpinnings of machine learning. The practical lab sessions will give learners an opportunity to implement machine learning algorithms and apply them to a range of datasets.

Assessment is divided into 2 elements. Firstly there are four worksheets which assess the learner's ability to understand and apply a machine learning algorithm to a dataset. Finally there is a written examination at the end of the semester which will assess the learner's theoretical understanding of the module.

Timetabling, learner effort and credit

The module is timetabled as one 1.5-hour lectures and one 1.5-hour lab per week.

The number of 5 ECTS credits assigned to this module is our assessment of the amount of learner effort required. Continuous assessment spreads the learner effort to focus on small machine algorithms and applications of these algorithms to suitable datasets.

There are 36 contact hours made up of 12 lectures delivered over 12 weeks with classes taking place in a classroom. There are also 12 lab sessions delivered over 12 weeks taking place in a fully equipped computer lab. The learner will need 57 hours of independent effort to further develop the skills and knowledge gained through the contact hours. An additional 32 hours are set aside for learners to work on worksheets that must be completed for the module.

The team believes that 125 hours of learner effort are required by learners to achieve the MIMLOs and justify the award of 5 ECTS credits at this stage of the programme.

Work-based learning and practice-placement

There is no work based learning or practice placement involved in the module.

E-learning

The college VLE is used to disseminate notes, advice, and online resources to support the learners. The learners are also given access to Lynda.com as a resource for reference.

Module physical resource requirements

Requirements are for a classroom for 60 learners equipped with a projector, and a 25 seater computer lab for practical sessions with access to Python development environments and associated libraries (this may change should more suitable technologies become available).

Reading lists and other information resources

Recommended Text

Géron, A. (2017) *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems*. Sebastopol: O'Reilly Media

Secondary Reading:

Goodfellow, I., et al., (2017) *Deep Learning*. Cambridge, MIT: Press

Specifications for module staffing requirements

For each instance of the module, one lecturer qualified to at least Bachelor of Science (Honours) in Computer Science or equivalent, and with a Certificate in Training and Education (30 ECTS at level 9 on the NFQ) or equivalent.. Industry experience would be a benefit but is not a requirement.

Learners also benefit from the support of the programme director, programme administrator, learner representative and the Student Union and Counselling Service.

Module Assessment Strategy

The assignments constitute the overall grade achieved, and are based on each individual learner's work. The continuous assessments provide for ongoing feedback to the learner and relates to the module curriculum.

No.	Description	MIMLOs	Weighting
1	Four worksheets each dealing with the implementation of a machine learning algorithm on an appropriate dataset	1 - 5	70%
2	Written exam which will test the theoretical aspects of the module.	1,2,3	30%

All repeat work is capped at 40%.

Sample assessment materials

Note: All assignment briefs are subject to change in order to maintain current content.

Course	BSCH
Stage / Year	4
Module	Machine Learning
Semester	2
Assignment	Worksheet 1
Date of Title Issue	x/x/x
Assignment Deadline	x/x/x (2 weeks after issue)
Assignment Submission	Upload to Moodle
Assignment Weighting	12.5% of module

- Each worksheet will follow a similar format to that presented below.
- If you have any questions, please address them to the following as appropriate:
 - course resources
 - online resources
 - partner
 - group
 - lecturer
- Please ensure that you can explain the following components of the lecture material to another student before proceeding with the worksheet:
 - technique and its parameters
 - dataset
 - *jupyter* notebook template
- All questions must be answered in the *jupyter* notebook
- Submission of your *jupyter* notebook must be made to the appropriate link on the module Moodle page.

Question 1 – 15 marks

- a) Describe the strengths and weaknesses of the following technique presented in this week's lecture: **Linear Regression**

Question 2 – 15 marks

- a) Summarize the following dataset presented in this week's lecture and identify 3 important question which could be answered by this dataset: **Passenger Numbers**

Question 3 – 40 marks

- a) Implement the technique using a *jupyter* notebook using the supplied template provided in the course resources as a guide.
b) Apply your implementation of the technique to the dataset.

Question 4 – 30 marks

- a) Describe the performance and accuracy of your implementation on the dataset in relation to the 3 questions you asked of the data.
b) Vary the 2 parameters of the technique and explain the change in performance and accuracy.
c) Explain where improvements can be made from a theoretical and implementation perspective. You may provide a summary of a variant of a technique that you have found online providing appropriate references.

Course	BSCH
Stage / Year	4
Module	Machine Learning
Semester	2
Assignment	Worksheet 2
Date of Title Issue	x/x/x
Assignment Deadline	x/x/x (2 weeks after issue)
Assignment Submission	Upload to Moodle
Assignment Weighting	12.5% of module

- Each worksheet will follow a similar format to that presented below.
- If you have any questions, please address them to the following as appropriate:
 - course resources
 - online resources
 - partner
 - group
 - lecturer
- Please ensure that you can explain the following components of the lecture material to another student before proceeding with the worksheet:
 - technique and its parameters
 - dataset
 - *jupyter* notebook template
- All questions must be answered in the *jupyter* notebook
- Submission of your *jupyter* notebook must be made to the appropriate link on the module Moodle page.

Question 1 – 15 marks

- a) Describe the strengths and weaknesses of the following technique presented in this week's lecture: **C4.5**

Question 2 – 15 marks

- a) Summarize the following dataset presented in this week's lecture and identify 3 important question which could be answered by this dataset: **Bank Details**

Question 3 – 40 marks

- a) Implement the technique using a *jupyter* notebook using the supplied template provided in the course resources as a guide.
b) Apply your implementation of the technique to the dataset.

Question 4 – 30 marks

- a) Describe the performance and accuracy of your implementation on the dataset in relation to the 3 questions you asked of the data.
b) Vary the 2 parameters of the technique and explain the change in performance and accuracy.
c) Explain where improvements can be made from a theoretical and implementation perspective. You may provide a summary of a variant of a technique that you have found online providing appropriate references.

Course	BSCH
Stage / Year	4
Module	Machine Learning
Semester	2
Assignment	Worksheet 3
Date of Title Issue	x/x/x
Assignment Deadline	x/x/x (2 weeks after issue)
Assignment Submission	Upload to Moodle
Assignment Weighting	12.5% of module

- Each worksheet will follow a similar format to that presented below.
- If you have any questions, please address them to the following as appropriate:
 - course resources
 - online resources
 - partner
 - group
 - lecturer
- Please ensure that you can explain the following components of the lecture material to another student before proceeding with the worksheet:
 - technique and its parameters
 - dataset
 - *jupyter* notebook template
- All questions must be answered in the *jupyter* notebook
- Submission of your *jupyter* notebook must be made to the appropriate link on the module Moodle page.

Question 1 – 15 marks

- a) Describe the strengths and weaknesses of the following technique presented in this week's lecture: **DBSCAN**

Question 2 – 15 marks

- a) Summarize the following dataset presented in this week's lecture and identify 3 important question which could be answered by this dataset: **Seattle Crime Data**

Question 3 – 40 marks

- a) Implement the technique using a *jupyter* notebook using the supplied template provided in the course resources as a guide.
b) Apply your implementation of the technique to the dataset.

Question 4 – 30 marks

- a) Describe the performance and accuracy of your implementation on the dataset in relation to the 3 questions you asked of the data.
b) Vary the 2 parameters of the technique and explain the change in performance and accuracy.
c) Explain where improvements can be made from a theoretical and implementation perspective. You may provide a summary of a variant of a technique that you have found online providing appropriate references.

Course	BSCH
Stage / Year	4
Module	Machine Learning
Semester	2
Assignment	Worksheet 4
Date of Title Issue	x/x/x
Assignment Deadline	x/x/x (2 weeks after issue)
Assignment Submission	Upload to Moodle
Assignment Weighting	12.5% of module

- Each worksheet will follow a similar format to that presented below.
- If you have any questions, please address them to the following as appropriate:
 - course resources
 - online resources
 - partner
 - group
 - lecturer
- Please ensure that you can explain the following components of the lecture material to another student before proceeding with the worksheet:
 - technique and its parameters
 - dataset
 - *jupyter* notebook template
- All questions must be answered in the *jupyter* notebook
- Submission of your *jupyter* notebook must be made to the appropriate link on the module Moodle page.

Question 1 – 15 marks

- a) Describe the strengths and weaknesses of the following technique presented in this week's lecture: **CNN**

Question 2 – 15 marks

- a) Summarize the following dataset presented in this week's lecture and identify 3 important question which could be answered by this dataset: **ImageNet**

Question 3 – 40 marks

- a) Implement the technique using a *jupyter* notebook using the supplied template provided in the course resources as a guide.
b) Apply your implementation of the technique to the dataset.

Question 4 – 30 marks

- a) Describe the performance and accuracy of your implementation on the dataset in relation to the 3 questions you asked of the data.
b) Vary the 2 parameters of the technique and explain the change in performance and accuracy.
c) Explain where improvements can be made from a theoretical and implementation perspective. You may provide a summary of a variant of a technique that you have found online providing appropriate references.

GRIFFITH COLLEGE DUBLIN

**QUALITY AND QUALIFICATIONS IRELAND
EXAMINATION**

SAMPLE PAPER

Machine Learning

Lecturer(s):

External Examiner(s):

Date: XXXXXXXX

Time: XXXXXXXX

**THIS PAPER CONSISTS OF FOUR QUESTIONS
FOUR QUESTIONS TO BE ATTEMPTED
ALL QUESTIONS CARRY EQUAL MARKS**

QUESTION 1

A database has ten transactions. Let minimum support be $S_{min} = 3$ or $S_{min} = 3/10 = 30\%$ and minimum confidence be $C_{min} = 80\%$.

TID	ITEMS
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

- (a) Define the notions of support and confidence, and using these metrics find all frequent items for the above database using the Apriori algorithm. Show your work.

(15 marks)

- (b) Compare the efficiency of the Apriori and FP Growth. Make a recommendation which algorithm to use with large real world datasets. Justify your choice.

(10 marks)

Total (25 marks)

QUESTION 2

Consider the training examples shown in the table below for a binary classification problem. We use the Decision Tree Induction algorithm to classify any new tuple based on the training examples given in the table below. The algorithm uses the *Information Gain* to select the test attributes.

CarID	Engine	Sc/Turbo	Weight	Fuel Eco	Fast
1	small	no	average	good	no
2	small	no	light	average	no
3	small	yes	average	bad	yes
4	medium	no	heavy	bad	yes
5	large	no	average	bad	yes
6	medium	no	light	bad	no
7	large	yes	heavy	bad	no
8	large	no	heavy	bad	no
9	medium	yes	light	bad	yes
10	large	no	average	bad	yes

11	small	no	light	good	no
12	small	no	average	average	no
13	medium	no	heavy	bad	no
14	small	yes	average	average	no
15	medium	no	heavy	bad	no

- (a) Compute the Information Gain necessary to classify any tuple of the examples given in the table above, and explain why CarID should not be used to achieve this even though it has the highest Information Gain.

(5 marks)

- (b) Compute the Information Gain for the Engine and SC/Turbo attributes.

(8 marks)

- (c) Compute the Information Gain for the Weight and Fuel Eco attributes.

(8 marks)

- (d) From your calculations which attribute provides the best information gain? Give a reason why..

(4 marks)

Total (25 marks)

QUESTION 3

- (a) With the aid of diagrams, explain the k-medoids algorithm.

(12 marks)

- (b) Explain the K-means algorithm.

(4 marks)

- (c) Compare K-mean, K-medoids, and DBSCAN in terms of the following criteria.

- (i) Shapes of the cluster that can be returned.

- (ii) Input parameters that must be specified.

(9 marks)

Total (25 marks)

QUESTION 4

- (a) With the aid of diagrams, discuss the suitability of a perceptron in modelling the following three logical operators: AND, OR, XOR.

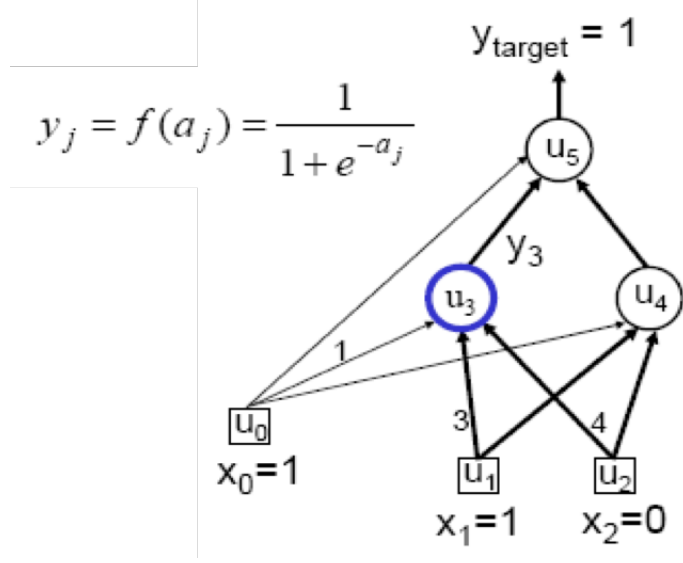
(9 marks)

- (b) With the aid of the Back Propagation algorithm discuss the forward pass of activity through a neural network including initialization.

(9 marks)

- (c) Based on the specified initialized weights of the network and the “squashing”

function of the neuron, determine the next output of the neuron u_3 .



(7 marks)

Total (25 marks)