Higher National Unit Specification

General information

Unit title: Machine Learning (SCQF level 7)

Unit code: J0J9 34

Superclass: CC

Publication date: August 2018

Source: Scottish Qualifications Authority

Version: 02

Unit purpose

The purpose of this unit is to introduce learners to the **foundations** of Machine Learning. This is a **non-specialist** unit. It is suitable for a wide range of learners but particularly those with an interest in STEM. No previous experience of programming or Artificial Intelligence is required.

The unit covers the **basic concepts** of Machine Learning. Learners will understand what Machine Learning is, the technological reasons for its growth, different approaches to Machine Learning, how Machine Learning is used, and the ethical issues surrounding its use in practice. The unit will help learners appreciate the applications and implications of this emerging technology.

At the completion of this unit, learners will understand the rationale for Machine Learning as an analytical tool and appreciate the applications of the technology in a variety of vocational fields. They will have the opportunity to explore contemporary developments in Machine Learning, such as deep learning.

Learners may progress to more advanced studies in Machine Learning or explore specific applications of Machine Learning. Appropriate Higher National units include SCQF level 8 *Machine Learning* which looks at Machine Learning programming, and DW8W 34 *Robotics and Animatronics: An Introduction*, which explores an application of Machine Learning.

Higher National Unit Specification: General information (cont)

Unit title: Machine Learning (SCQF level 7)

Outcomes

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

1 Explain the fundamentals of Machine Learning.
2 Describe the components of a Machine Learning system.
3 Explain different learning approaches used in Machine Learning.
4 Develop models for prediction and clustering using a Machine Learning workbench.

Credit points and level

1 Higher National unit credit at SCQF level 7: (8 SCQF credit points at SCQF level 7)

Recommended entry to the unit

While entry is at the discretion of the centre, it would be beneficial if learners possessed numeracy skills, which may be evidenced by possession of the Core Skills unit in *Numeracy* at SCQF level 6 (F3GF 12). Some previous knowledge of statistics would be desirable but not essential. It would also be beneficial if learners possessed IT skills, which may be evidenced by possession of the Core Skills unit in *Information and Communication Technology* at SCQF level 6 (F3GC 12).

Core Skills

Achievement of this Unit gives automatic certification of the following Core Skills component:

Complete Core Skill	None
Core Skill component	Critical Thinking at SCQF level 5

Context for delivery

If this unit is delivered as part of a group award, it is recommended that it should be taught and assessed within the subject area of the group award to which it contributes.

Equality and inclusion

This unit specification has been designed to ensure that there are no unnecessary barriers to learning or assessment. The individual needs of learners should be taken into account when planning learning experiences, selecting assessment methods or considering alternative evidence.

Further advice can be found on our website **www.sqa.org.uk/assessmentarrangements**.

Higher National Unit Specification: Statement of standards

Unit title: Machine Learning (SCQF level 7)

Acceptable performance in this unit will be the satisfactory achievement of the standards set out in this part of the unit specification. All sections of the statement of standards are mandatory and cannot be altered without reference to SQA.

Where evidence for outcomes is assessed on a sample basis, the whole of the content listed in the knowledge and/or skills section must be taught and available for assessment. Learners should not know in advance the items on which they will be assessed and different items should be sampled on each assessment occasion.

Outcome 1

Explain the fundamentals of Machine Learning.

Knowledge and/or skills

- Definitions of Artificial Intelligence, Machine Learning and deep learning
- Relationship of Machine Learning to other approaches in the field of Artificial Intelligence
- Neural networks and deep learning
- Technological reasons for the growth of Machine Learning and its historical development
- Difference between Machine Learning and statistical approaches to modelling
- Applications of Machine Learning in business, science and society
- Ethical issues raised by the use of Machine Learning including algorithmic bias
- Machine Learning cycle

Outcome 2

Describe the components of a Machine Learning system.

Knowledge and/or skills

- Algorithms, including clustering, decision tree and Bayesian algorithms
- Structured and unstructured data sources and acquisition
- Data cleansing
- Computational methods and algorithms
- Training and testing data sets
- Model development
- Feature selection and dimensionality reduction
- Evaluation of Machine Learning outcomes

Higher National Unit Specification: Statement of standards (cont)

Unit title: Machine Learning (SCQF level 7)

Outcome 3

Explain different learning approaches used in Machine Learning.

Knowledge and/or skills

- Supervised learning and its use in prediction
- Common algorithms for supervised learning
- Unsupervised learning and its use in clustering
- Common algorithms for unsupervised learning
- Reinforcement learning and its use in maximising system performance

Outcome 4

Develop models for prediction and clustering using a Machine Learning workbench.

Knowledge and/or skills

- Programming languages in use for Machine Learning computation
- Machine Learning workbench as an alternative to coding
- Application of classification algorithm in a Machine Learning workbench
- Application of clustering algorithm in a Machine Learning workbench

Evidence requirements for this unit

Learners will need to provide evidence to demonstrate the knowledge and/or skills across all outcomes. The evidence requirements for this unit will take **two** forms.

1 Knowledge evidence
2 Product evidence

The **knowledge evidence** will relate to all outcomes. Knowledge evidence is required for all knowledge and/or skills statements of Outcomes 1, 2 and 3, as well as underpinning knowledge in practical outcomes (Outcome 4). The evidence may be produced over an extended period of time in lightly controlled conditions. The amount of evidence may be the **minimum** required to infer competence. For example, it is acceptable if learners only describe the most important ethical issues relating to Machine Learning (Outcome 1) and provide basic explanations of supervised learning (Outcome 3).

The knowledge evidence may be sampled when testing is used. In this case, the evidence must be produced under controlled conditions in terms of location (supervised), timing (limited) and access to reference materials (not permitted). The sampling frame must cover all outcomes but not all knowledge/skills statements; however, the majority of the knowledge/skills should be sampled in every test. The sampling frame must always include the following:

- definition of Machine Learning
- technological reasons for the growth of Machine Learning and its historical development
- current applications of Machine Learning in business, society and science
- ethical issues raised by the use of Machine Learning
- algorithms
- training and testing data sets
- evaluation of Machine Learning outcomes

The knowledge evidence may be written or oral, or a combination of these. Evidence may be captured, stored and presented in a range of media (including audio and video) and formats (analogue and digital). Particular consideration should be given to digital formats and the use of multimedia.

The **product evidence** will relate to Outcome 4. The evidence must demonstrate that the learner can apply **at least one** packaged method to **at least one** given data set to produce **at least one** classification or clustering model and an estimate of its performance. This evidence may be produced over the life of the unit, under loosely controlled conditions (including access to reference materials). Authentication will be necessary (see below).

The SCQF level of this unit (level 7) provides additional context on the nature of the required evidence and the associated standards. Appropriate level descriptors should be used when making judgements about the evidence.

When evidence is produced in loosely controlled conditions, it must be authenticated. The Guide to Assessment provides further advice on methods of authentication.

The support notes section of this specification provides specific examples of instruments of assessment that will generate the required evidence.

Higher National Unit Support Notes

Unit title: Machine Learning (SCQF level 7)

Unit support notes are offered as guidance and are not mandatory.

While the exact time allocated to this unit is at the discretion of the centre, the notional design length is 40 hours.

Guidance on the content and context for this unit

Machine Learning has taken on a particular significance in the modern world. Most are aware of the influence that Machine Learning has on their personal life, from credit application scoring to tailored recommendations from online retailers and media providers, and the ethical questions this raises.

Therefore, the unit is intended for anyone with an interest to understand the basic concepts of Machine Learning — how it works, the tools it deploys, and the reliability of the decisions it supports.

The treatment of algorithms and model acquisition will be simple and straightforward, but sufficient to provide an insight into the mechanistic nature of Machine Learning. The unit will provide an overview of Machine Learning as one facet of Artificial Intelligence, and its relationship to other approaches in this field. It will also introduce the concepts of error and bias in model development, and their importance in evaluating model performance. The unit will introduce learners to the use of packaged tools to perform a few common Machine Learning tasks.

The ubiquitous nature of Machine Learning deployment in the modern world suggests that tailoring the content to specific learner contexts (eg, such as business) should be avoided to allow the range of ethical issues to be explored. It will be important to present the applications and implications of Machine Learning in a balanced way, neither overstating its potential nor understating the risks that it also poses.

Please note that the following guidance, relating to specific outcomes, does not seek to explain each knowledge/skills statement, which is left to the professionalism of the teacher. It seeks to clarify the statement of standards where it is potentially ambiguous. It also focuses on non-apparent teaching and learning issues that may be over-looked, or not emphasised, during unit delivery. As such, it is not representative of the relative importance of each knowledge/skill.

Outcome 1

This outcome relates to the contexts in which Machine Learning has developed, and its relationship to other aspects of Artificial Intelligence, such as deep learning and neural networks. The increasing availability of big data and the technological advances in computing power and storage should also be explained as contributors to the increasing ability to perform the complex computation associated with Machine Learning algorithms.

The knowledge statements are largely self-explanatory, but the treatment of these concepts should be relatively light, and sufficient to understand the contexts in which the basic ideas of Machine Learning will be explored in the other outcomes. The mechanistic nature of Machine Learning should be stressed, with simple examples introduced to give an understanding of how models work (eg, decision trees and linear models). The comparison of Machine Learning with statistical modelling should be taken as an opportunity to introduce basic concepts of descriptive statistics and inferential statistics, along with basics of probability to explain Bayes Theorem.

In presenting applications of Machine Learning in business, society and science, there should be a focus on aspects of data provenance and quality and the ethical issues raised by the application of models developed in this way. The risks of bias and error should be briefly covered.

Outcome 2

In Outcome 2, the importance of knowing and understanding the source of a data set will be further developed from Outcome 1, along with an overview of where and how data sets can be acquired for learning and practice purposes (eg, from Twitter, sales data, UCL database). The consideration of data cleansing should be limited to the treatment of missing data and inaccurate data (eg, misspelled names, out-of-range values). The concept of data normalisation prior to its use in Machine Learning should be given light coverage.

In this outcome, the basic concept of algorithms should be introduced, along with the role that computation plays in their application, particularly to large sets of data. The mechanistic nature of computation should be stressed. The purpose of training and testing data sets is an important concept in this unit, and the role that they play in model development and performance measuring. There should be a light treatment of the concept that some features in the data may contribute little to the ability of a model to classify, predict or cluster. Removing these features can reduce the complexity and size of the data set without harming the performance of the developed model. This could be demonstrated by use of a Machine Learning workbench (such as Weka).

Learners should be made aware of the basic measures of performance that can be used in model development, covering classification, prediction and clustering.

Outcome 3

Outcome 3 provides the opportunity to engage the learner directly with the core concepts of Machine Learning. Simple examples (eg, two or three dimensional) will be sufficient to demonstrate the concepts of classification, prediction and clustering.

The differences between supervised, unsupervised and reinforcement learning should be introduced, with the emphasis mainly on the first two. The range of algorithms covered in this outcome should be constrained to two or three common algorithms that would be understood in the context of this non-specialist module. Only algorithms for classification, prediction and clustering should be covered, and only for supervised and unsupervised learning.

An appropriate treatment of this topic would be to demonstrate algorithm performance using pre-packaged algorithms in a programmatic environment (eg, Python with scikit-learn) or in a Machine Learning workbench (eg, Weka). Learners should gain an appreciation of the role that the algorithm plays in the production of a model from the given data. It is suggested that this should be constrained to a single example such as a classification problem.

Outcome 4

In Outcomes 2 and 3 there will be opportunities to provide practical demonstration of Machine Learning through use of a programmatic environment or a Machine Learning workbench. Learners should know that there are various programmatic environments for coding Machine Learning algorithms, and the concept of packaged solutions in these environments. Python (with scikit-learn), R Studio (with class, tree, c50), and MATLAB would merit a mention.

The concept of a Machine Learning workbench should be introduced, focusing on their ability to support non-programmers to deploy computational algorithms for the purposes of Machine Learning. Learners should be made aware that IBM, Google, Apple and Microsoft Azure all offer access to huge amounts of computing power in the cloud. A light treatment of how users would access these services, along with a demonstration, should suffice.

However, learners will have to be given sufficient familiarity with one Machine Learning workbench (such as the open source package Weka) to enable them to demonstrate skills in the development of a model for supervised learning for a given scenario. The data sets used for the scenario should have a reasonably small number of features, but have a sufficiently large number of items to allow conclusions to be drawn about model performance. It is suggested that a classification problem or a clustering problem would work best at this level.

This outcome will cover the application of the workbench to the given data set and drawing conclusions about the performance of the derived model.

Guidance on approaches to delivery of this unit

A suggested distribution of time, across the outcomes, is:

Outcome 1: 8 hours
Outcome 2: 10 hours
Outcome 3: 10 hours
Outcome 4: 12 hours

Summative assessment may be carried out at any time. However, when testing is used (see evidence requirements) it is recommended that this is carried out towards the end of the unit (but with sufficient time for remediation and re-assessment). When continuous assessment is used (such as the use of a web log), this could commence early in the life of the unit and be carried out throughout the duration of the unit.

There are opportunities to carry out formative assessment at various stages in the unit. For example, formative assessment could be carried out on the completion of each outcome to ensure that learners have grasped the knowledge contained within it. This would provide assessors with an opportunity to diagnose misconceptions, and intervene to remedy them before progressing to the next outcome.

Higher National Unit Support Notes (cont)

Unit title: Machine Learning (SCQF level 7)

Guidance on approaches to assessment of this unit

Evidence can be generated using different types of assessment. The following are suggestions only. There may be other methods that would be more suitable to learners.

Centres are reminded that prior verification of centre-devised assessments would help to ensure that the national standard is being met. Where learners experience a range of assessment methods, this helps them to develop different skills that should be transferable to work or further and higher education.

The assessment for this unit will require evidence of knowledge and understanding of the concepts introduced in each of the four outcomes of this unit, along with practical evidence of the skills stated in Outcome 4. Assessment of knowledge and understanding could be generated from an end of unit test, or from a portfolio developed over the course of the unit. Each is described in more detail below.

The end of unit test would sample from the knowledge and understanding contained in all four outcomes. The test could comprise a number of Extended Response Questions (ERQs) and would be marked and assessed traditionally. For example, the test may comprise of five extended response questions each worth 10 marks, with the learner responses marked out of 50, with a pass mark of 25. If this approach is taken, it is recommended that some (or all) of the ERQs combine the knowledge and understanding within and across outcomes. This test would be taken, sight-unseen, in controlled and timed conditions without reference to teaching materials. A suitable duration could be 90 minutes.

A more contemporary approach to assessment would involve the use of an e-portfolio (such as Mahara) to record learning (and its associated activities) throughout the life of the unit. The e-portfolio would provide evidence of knowledge and understanding through descriptions and explanations. The e-portfolio should be assessed using defined criteria to permit a correct assessment judgement to be reached about the quality of the evidence. In this approach, every knowledge and skill must be evidenced; sampling would not be appropriate.

Formative assessment could be used to assess learners' knowledge at various stages throughout the life of the unit. An ideal time to gauge the extent of knowledge and understanding would be at the end of each outcome. Formative assessment could be delivered through an item bank of selected response questions, providing diagnostic feedback to learners (when appropriate).

If an e-portfolio is used for summative assessment, it would also facilitate formative assessment since learning (including misconceptions) would be apparent from the e-portfolio, and intervention could take place to correct misunderstandings on an on-going basis.

The practical evidence will be in the form of product evidence, comprising a short report for a data set supplied by the centre. The report will detail the Machine Learning algorithm used, the justification for its use, the software environment in which Machine Learning took place, the generated model and an evaluation of its performance. The evidence will be assessed using a set of defined criteria.

Higher National Unit Support Notes (cont)

Unit title: Machine Learning (SCQF level 7)

Opportunities for e-assessment

E-assessment may be appropriate for some assessments in this unit. By e-assessment we mean assessment which is supported by Information and Communication Technology (ICT), such as e-testing or the use of e-portfolios or social software. Centres which wish to use e-assessment must ensure that the national standard is applied to all learner evidence and that conditions of assessment as specified in the evidence requirements are met, regardless of the mode of gathering evidence. The most up-to-date guidance on the use of e-assessment to support SQA's qualifications is available at **www.sqa.org.uk/e-assessment**.

Opportunities for developing Core and other essential skills

Opportunities to develop aspects of Core Skills occur in Outcomes 2, 3 and 4. These are:

- I*nformation and Communication Technology (ICT)* at SCQF level 6
- *Numeracy* at SCQF Level 6

In terms of *Information and Communication Technology*, there are opportunities in the use of software in this unit to demonstrate evidence of starting software, entering and editing data, locating and extracting information and evaluating and presenting information. There will be directed learning to seek out examples of the application of Machine Learning in business, science and society that will require the application of a complex search strategy and the evaluation and presentation of information.

In relation to *Numeracy*, there are opportunities to identify relevant data and relationships, use statistical concepts and identify significant features in complex graphical information. The unit will also provide an opportunity to develop broader skills, such as citizenship, when considering the ethical implications of Machine Learning (Outcome 1), along with analytical and reasoning skills when evaluating a Machine Learning model (Outcome 4).

This Unit has the Critical Thinking component of Problem Solving embedded in it. This means that when learners achieve the Unit, their Core Skills profile will also be updated to show they have achieved Critical Thinking at SCQF level 5.

History of changes to unit

Version	Description of change	Date
02	Core Skills Component Critical Thinking at SCQF level 5 embedded.	31/08/18

General information for learners

Unit title: Machine Learning (SCQF level 7)

This section will help you decide whether this is the unit for you by explaining what the unit is about, what you should know or be able to do before you start, what you will need to do during the unit and opportunities for further learning and employment.

The purpose of this unit is to provide you with an introduction to the theory and practice of Machine Learning. It is intended for anyone who has an appreciation of the importance of Machine Learning to their personal and professional life and wish to understand better the fundamental concepts on which its application is based.

The unit comprises four outcomes and develops your knowledge and understanding of the concepts behind Machine Learning along with some practical competence in using software tools that perform Machine Learning.

Outcome 1 provides you with an overview of Machine Learning and explains its relationship to other Artificial Intelligence methods and the main drivers for its development and rapid expansion. You will have the opportunity to identify and explore practical applications from business, science (including medicine) and social contexts, and through these to consider the ethical challenges that the use of Machine Learning brings.

Outcome 2 provides you with an introduction to the concept of a Machine Learning workflow, and the importance of understanding the decisions that are taken at each stage of this process. These include the identification of relevant data sets, their provenance and reliability, the need to reduce dimensionality and normalise data. The workflow includes the training and testing of data sets.

Outcome 3 introduces you to the differences between supervised, unsupervised and reinforcement learning, with the emphasis mainly on the first two. In each case, you will only cover two or three common algorithms, to keep within the non-specialist nature of this unit. You will experience demonstrations of algorithm performance using pre-packaged algorithms in a programmatic environment (eg, Python with scikit-learn) or in a Machine Learning workbench (eg, Weka).

Outcome 4 introduces you to a sample of programmatic environments for coding Machine Learning algorithms, and the concept of packaged solutions in these environments. Typical of these are Python (with scikit-learn), R Studio (with class, tree, c50), and MATLAB on. You will learn the concept of a Machine Learning workbench and become aware of the availability of cloud-based Machine Learning services from IBM, Google, Apple and Microsoft.

The unit requires you to undertake practical work to develop a model for supervised learning for a given scenario, and report on your findings, using a data set supplied by the centre.

This unit will provide you with the underpinning knowledge and skills to progress to the SCQF level 8 unit in *Machine Learning* It will also afford opportunity for you to enhance your competence in Core Skills such as *ICT*, *Numeracy* and statistics.

This Unit has the Critical Thinking component of Problem Solving embedded in it. This means that when you achieve the Unit, your Core Skills profile will also be updated to show you have achieved Critical Thinking at SCQF level 5.