Next Generation Higher National Unit Specification

Machine Learning (SCQF level 8)

Unit code: J6CN 48

SCQF level: 8 (16 SCQF credit points)

Valid from: session 2023–24

Prototype unit specification for use in pilot delivery only (version 1.0) September 2023

This unit specification provides detailed information about the unit to ensure consistent and transparent assessment year on year.

This unit specification is for teachers and lecturers and contains all the mandatory information required to deliver and assess the unit.

This edition: September 2023 (version 1.0)

Unit purpose

This specialist unit introduces learners to practical machine learning. It is suitable for those with an interest in computer science, statistics or artificial intelligence (AI). Previous experience of computer programming and well-developed numeracy skills are required. We recommend that learners complete the Machine Learning unit at SCQF level 7 before doing this unit.

Learners gain an understanding of how machine learning algorithms can be implemented in an appropriate programming language. While they don't need previous experience of specific languages, learners should have previous programming experience in a high-level language and be familiar with programming concepts and techniques.

They gain knowledge and understanding of the syntax and semantics of a specific machine learning language, and use that understanding to investigate a range of common machine learning approaches and their associated algorithms. Learners understand the importance of well-behaved data and the approaches to selecting, cleaning and transforming datasets. They learn how to evaluate model performance and develop models that minimise bias and error.

On completing this unit, learners can progress to more advanced studies in machine learning, including more advanced programming techniques.

Unit outcomes

Learners who complete this unit can:

1 explain the characteristics of different machine learning methods and models
2 distinguish between various types of machine learning algorithms
3 implement programme code to transform data and apply algorithms
4 build machine learning models, including improvement of model performance

Evidence requirements

Learners must provide both knowledge and product evidence.

Knowledge evidence

Evidence can be sampled when testing is used. In this case, learners must produce evidence under controlled conditions in terms of location (supervised), timing (limited) and access to reference materials (not allowed). Sampling must include outcomes 1, 2 and 4 (at least partially) and most knowledge and skills statements. The sample must always include the following:

- identify supervised, unsupervised and reinforcement learning methods and explain their differences
- explain the basic principle of predictive modelling
- identify and allocate machine learning models to a category: classification, regression or clustering
- identify and explain how the most common algorithms operate for each of these purposes:
  - classification
  - regression
  - clustering
- identify and explain the basic algorithms associated with instance-based machine learning
- describe approaches to the reduction of dimensionality in data
- identify and explain approaches to training and validation in machine learning, including cross-validation
- describe approaches to feature selection and transformation in machine learning
- explain the concept of overfitting in machine learning
- identify and explain the metrics associated with machine learning performance

Evidence can be written or oral or a combination of these. It can be captured, stored and presented in a range of media (including audio and video) and formats (analogue and digital).

Product evidence

This evidence relates to outcomes 3 and 4. It demonstrates that learners have the competence to write and implement program code to manage data and apply selected algorithms using program packages. It also shows that learners have the skills to build at least one machine learning model for each of the following scenarios:

- classification
- clustering
- regression

Product evidence also demonstrates the skill of applying methods to improve the machine learning model using ensemble approaches.

Learners can produce this evidence throughout the unit, under lightly-controlled conditions (including access to reference materials) and must be authenticated. The Guide to Assessment provides further advice on methods of authentication.

The standard of evidence should be consistent with the SCQF level of this unit.

You should use appropriate level descriptors when making judgements about the evidence.

Knowledge and skills

The following table shows the knowledge and skills covered by the unit outcomes.

Knowledge	Skills
Learners should understand: ♦ supervised learning ♦ unsupervised learning ♦ reinforcement learning ♦ ensemble learning (improving performance) ♦ predictive modelling (function approximation) ♦ classification models ♦ regression models ♦ clustering models ♦ classification algorithms ♦ regression algorithms ♦ clustering algorithms ♦ instance-based algorithms ♦ dimensionality reduction ♦ programming systems for machine learning ♦ machine learning workflow ♦ training and validation ♦ cross-validation ♦ feature transformation and selection ♦ bias, error and precision-recall metrics ♦ overfitting	Learners can: ♦ clean and normalise data ♦ visualise data ♦ write program code ♦ use program packages ♦ perform coding to implement machine learning algorithms ♦ build machine learning models ♦ perform ensemble learning to improve model performance

Meta-skills

Throughout the unit, learners develop meta-skills to enhance their employability in the data science sector.

Self-management

This meta-skill includes:

- focusing: sorting, attention, filtering
- adapting: openness, critical reflection, adaptability, self-learning
- initiative: independent thinking, decision making

Social intelligence

This meta-skill includes:

- communicating: receiving information, giving information, storytelling
- feeling: social conscience

Innovation

This meta-skill includes:

- curiosity: observation, questioning, information sourcing, problem recognition
- creativity: idea generation, visualising, maker mentality
- sense-making: pattern recognition, holistic thinking, synthesis, opportunity recognition, analysis
- critical thinking: deconstruction, logical thinking, judgement, computational thinking

Literacies

Throughout this unit, learners have opportunities to develop their literacy skills.

Numeracy

Learners develop numeracy skills by using different approaches to manipulate data.

Communication

Learners develop communication skills by using data visualisation to explore and understand large data sets.

Digital

Learners develop digital skills and computer literacy throughout this unit.

Delivery of unit

If you deliver this unit as part of a group award, we recommend that you teach and assess it in the subject area of that group award.

By taking a holistic approach, learners develop understanding and skills together. In particular, they should gain the knowledge and skills in outcomes 1, 2 and 3 by studying particular models and their algorithms, along with their implementation on practice datasets. Outcome 4 deals with the final stage of performance evaluation and prediction improvement, but you can also deliver this in the context of each model and algorithm combination.

You must teach all of the content listed in the 'Knowledge and skills' section, even if you assess evidence for outcomes on a sample basis. Learners should not know in advance the items being assessed, and you should sample different items on each assessment occasion.

While the exact time allocated to this unit is at your centre's discretion, the notional design length is 80 hours.

We suggest the following distribution of time:

Outcome 1 — Explain the characteristics of different machine learning methods and models (15 hours)

Outcome 2 — Distinguish between various types of machine learning algorithms (20 hours)

Outcome 3 — Implement programme code to transform data and apply algorithms (25 hours)

Outcome 4 — Build machine learning models, including improvement of model performance (20 hours)

Additional guidance

The guidance in this section is not mandatory.

Recommended entry to the unit

While entry is at your centre's discretion, we recommend that learners have a basic knowledge of the underpinning concepts of machine learning, along with some practical skills in computer programming. Evidence of this can be if learners have completed the Machine Learning unit at SCQF level 7. Alternatively, completing the Artificial Intelligence unit at SCQF level 7 or the Big Data unit at SCQF level 7 can provide sufficient underpinning knowledge and understanding.

Learners can provide evidence of programming by completing a unit such as the Software Development unit at SCQF level 7. A basic understanding of statistical concepts would be beneficial, but is not essential.

Content and context for this unit

Learners gain confidence and competence in selecting and applying machine learning algorithms to build models for prediction, classification or clustering. They develop a familiarity with a chosen programming environment and the machine learning packages that it supports (such as Python, R or MATLAB).

The following guidance clarifies the statement of standards, and focuses on non-apparent learning and teaching issues that could be over-looked or not emphasised when delivering the unit. As such, it is not representative of the relative importance of each knowledge and skill — explaining each knowledge and skills statement is at your discretion.

At the time of writing, this unit does not lead to recognition by a professional body. It does however provide some underpinning knowledge for the Data Science unit at SCQF level 8.

Explain the characteristics of different machine learning methods and models (outcome 1)

This builds on previous knowledge of the concept of machine learning and provides a broad overview of the main models and approaches that learners encounter in developing a machine learning solution for a dataset. The main learning points are:

- predictive modelling can be represented as function approximation, as exemplified by classification, regression and clustering
- ensemble learning employs a range of techniques to reduce bias and variance in prediction

Distinguish between various types of machine learning algorithms (outcome 2)

This focuses on the range of algorithms that can be deployed in machine learning, and their relationship to the methods and models covered in outcome 1. This should not be an exhaustive list of algorithms, but you should present the following as the core algorithms for machine learning:

- Decision trees and their related ID3 algorithm (information gain metric), followed by support vector machine (SVM) and then Naïve Bayes to introduce the probabilistic element.
- Prediction of continuous variables through regression approaches, including linear and polynomial regression (least squares), followed by logistic regression to link to the classification model.
- Clustering algorithms: these can be limited to K-means, Hierarchical and Gaussian along with the metrics used to judge their nearest neighbour.
- Instance-based algorithms, limited to the K-NN classification algorithm.
- Algorithms for dimensionality reduction, introduced as:
  - selection (variance and correlation thresholds)
  - extraction (principal components analysis)

Implement programme code to transform data and apply algorithms (outcome 3)

Introduce learners to a programming environment that supports machine learning packages for rapid implementation of machine learning in code. Typically these are: Python (with scikit-learn); MATLAB (with machine learning toolbox); and R (with caret, tree and e1071). These environments are continually expanded or upgraded to encompass new features (for example R now has the cloudml package to use Google Cloud Machine Learning Workbench), so you should pay attention to choosing the best environment for any target group.

You should cover the data cleaning functionality in the chosen programming environment (such as replacing missing values and identifying out-of-range values) as well as packages for data visualisation such as ggplot2 (R), matplotlib (Python), or the MATLAB Visualizations app. Workbenches, such as the Microsoft Azure ML Studio, also support data visualisation. This outcome includes a number of practical projects that contribute to assessing the overall unit.

Build machine learning models, including improvement of model performance (outcome 4)

An important part of developing a machine learning model is the process of training and validation. You should explain their purpose and examine the relationship with cross-validation (Holdout and K-fold cross-validation). You should also cover the application of methods for feature transformation and selection. Learners should understand the importance of bias and error in model-fitting, and the role of the precision-recall metric in judging model performance. You should provide examples of their relationship with overfitting, leading on to the importance of ensemble methods. The practical work requires the fitting of a model, and the application of bagging and/or boosting ensemble methods (such as Random Forest or ADABoost) to improve the prediction from a fitted model.

The popularity of data science has resulted in a wide range of online resources for learners wanting to gain a deeper understanding of machine learning. These range from exemplar datasets to tutorials on specific algorithms and programming language implementations. They also include complete online courses from providers, such as Coursera, edX and FutureLearn, and extensive learning materials from machine learning workbench providers, such as Microsoft, Google and Apple.

Useful online resources include:

- Data Science Solutions (OCI)
- Elite Data Science
- Machine Learning Mastery

Datasets for practical work can be found on the UCI database and also on kaggle.

Animated visualisation is a powerful method for understanding how different machine learning algorithms work. There are many online resources that can assist this approach to unit delivery, such as can be found at R2D3.

Approaches to assessment

Evidence can be generated using different types of assessment. The following are suggestions, however, there may be other methods that could be more suitable for your learners.

Assessment could be carried out using:

- an assignment that covers the knowledge and understanding for outcome 2
- a selected-response test that covers the knowledge and understanding for outcomes 1 and 4
- a set of practical tasks that cover the practical competence and understanding for outcomes 3 and 4

The assignment could ask learners to research and present evidence describing the various types of machine learning algorithms, including their structure and purpose. The evidence must cover knowledge of classification, clustering, regression and instance-based algorithms. This should be in their own words, and include references to information sources. This assignment would be carried out over a set period of time.

You could structure each selected-response question as four options (one correct answer and three distractors), with a 60 per cent pass mark for the whole test. Use scenario-type questions to assess competence in distinguishing methods and models. The test could include a high number of questions (for example 30 or 40), last 1 hour, cover outcomes 1 and 4, and sample all of the knowledge statements (including at least one question for each statement).

A set of practical tasks could be carried out over an extended period of time. This would allow learners to demonstrate competence implementing programme code to transform data, and apply algorithms to problem data sets. They could also show competence in improving a fitted model as part of the practical tasks, however, the tasks must cover all of the practical competences set out in outcomes 3 and 4.

A more contemporary approach to assessment would be learners using a blog to record learning (and the associated activities) throughout the unit. The blog could provide knowledge evidence (in the descriptions and explanations) and performance evidence (using, for example, computing logs). You should assess the blog using defined criteria to judge the quality of the digital evidence. In this approach to assessment, there must be evidence of every knowledge and skill — sampling is not sufficient.

You can use formative assessment to assess knowledge at various stages throughout the unit. An ideal time is at the end of each outcome. You can deliver the assessment through an item bank of selected-response questions, providing diagnostic feedback to learners (where appropriate).

If learners use a blog for summative assessment, it also helps formative assessment, as learning (including misconceptions) is apparent from the blog, and you can correct misunderstandings on an on-going basis.

You can carry out summative assessment at any time. However, when you use testing (see the ‘Evidence requirements’ section), we recommend that you do so towards the end of the unit.

If you use continuous assessment (such as using a blog), you can start this early in the unit and continue throughout.

You have opportunities to carry out formative assessment at various stages in the unit. For example, on completing each outcome, to ensure that learners have grasped the knowledge contained within it. This provides you with an opportunity to diagnose misconceptions and intervene to remedy them before progressing to the next outcome.

Equality and inclusion

This unit is designed to be as fair and as accessible as possible with no unnecessary barriers to learning or assessment.

You should take into account the needs of individual learners when planning learning experiences, selecting assessment methods or considering alternative evidence.

Guidance on assessment arrangements for disabled learners and/or those with additional support needs is available on the assessment arrangements web page.

Information for learners

Machine Learning (SCQF level 8)

This information explains:

- what the unit is about
- what you should know or be able to do before you start
- what you need to do during the unit
- opportunities for further learning and employment

Unit information

This unit enhances your knowledge of current approaches to machine learning and develop your competence in using software to implement common algorithms used in machine learning. You should have prior experience of machine learning approaches and have completed the Machine Learning unit at SCQF level 7 or equivalent.

You learn about supervised, unsupervised and reinforcement machine learning methods, and gain an understanding of how they can be combined in ensemble learning. The unit also explains the range of models that you might encounter in machine learning — predictive, classification, regression and clustering.

You are introduced to the algorithms that relate to each of these models, including decision trees, linear and logistic regression, K-means and hierarchical clustering, and K-NN classification. You also learn how these algorithms can be implemented through existing procedure libraries.

An important element of machine learning lies in preparing data. You learn about approaches to cleaning and normalising data, and how to use data visualisation packages to explore and understand data. You also learn about reducing the dimensionality of large data sets using approaches such as extraction, selection and principal axis factoring (PAF).

You practise machine learning by implementing algorithms in code to prepare data, applying a model, and judging the quality of the learned model. To do this, you develop competence in using a selected programming environment that supports machine learning, such as Python, R or MATLAB. In your chosen environment, you become familiar with the libraries of coded routines, that complete the various tasks that comprise the machine learning workflow.

You learn how to build a machine learning model and then improve its performance. You develop an understanding of how to use training and validation subsets of your data, and the role of cross-validation.

Finally, you learn the importance of and how to use the various metrics of model performance — such as bias, error and precision-recall — and how ensemble learning can improve model performance. The dangers of overfitting are also covered.

During assessment, you can provide evidence of your knowledge and understanding of the characteristics of the various machine learning methods and models, and also of the types of algorithms commonly used. You can show your knowledge and understanding of the machine learning workflow and the role of data manipulation and transformation, as well as the measures used to judge model performance. The practical assignments allow you to demonstrate your competence in developing code solutions that perform machine learning.

You develop meta-skills in self-management, social intelligence and innovation.

On completing this unit, you can progress to more advanced studies in machine learning, including more advanced programming techniques, such as the Machine Learning unit at SCQF level 9.

Administrative information

Published: September 2023 (version 1.0)

Superclass: CB

History of changes

Version	Description of change	Date

Note: please check SQA's website to ensure you are using the most up-to-date version of this document.