National Unit Specification

General information

Unit title: Machine Learning (SCQF level 6)

Unit code: J891 46

Superclass: CB

Publication date: June 2024

Source: Scottish Qualifications Authority

Version: 01

Unit purpose

The purpose of this non-specialist unit is to develop understanding of the concepts that underpin machine learning and the methods used to implement it. It provides a practical introduction that will develop basic skills in training and evaluating common machine learning models. Implementing the machine learning workflow will provide experience of data preprocessing, feature selection, model training and evaluation of model fit.

Entry to this unit is at the discretion of your centre. It would be advantageous for learners to have a basic understanding of data science concepts and be familiar with computer operation and software use. This may be evidenced by possession of one or more units from the NPA in Data Science at SCQF level 5 or higher.

On completion of this unit, learners will have a sound understanding of the basic concepts of machine learning, its benefits and challenges, and be able to independently implement machine learning workflows to solve problems with structured data.

Learners may wish to progress to further study in J0J9 34 Machine Learning (SCQF level 7). They may also apply their knowledge and skills to a data science project, such as J2GT 46 Data Science Project (SCQF level 6).

National Unit Specification: General information (continued)

Unit title: Machine Learning (SCQF level 6)

Outcomes

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

1. Demonstrate a conceptual understanding of machine learning and its applications.
2. Explain the role of algorithms in supervised machine learning.
3. Explain how the machine learning workflow contributes to model development.
4. Apply the machine learning workflow to train and evaluate supervised learning models.

Credit points and level

1 National Unit credit at Scottish Credit and Qualifications Framework (SCQF) level 6: (6 SCQF credit points at SCQF level 6).

Recommended entry to the unit

Entry is at the discretion of the centre. No prior knowledge of machine learning is required, although learners should have familiarity with computational methods and skills in numeracy.

Core Skills

Opportunities to develop aspects of Core Skills are highlighted in the support notes for this unit specification.

There is no automatic certification of Core Skills or Core Skill components in this unit.

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.

For example, if this unit is delivered as part of the National Progression Award in Data Science at SCQF level 6 there is overlap with other units within this award (particularly Data Science at SCQF level 6) and there will be opportunities to contextualise and integrate teaching, learning and assessment across component units.

National Unit Specification: General information (continued)

Unit title: Machine Learning (SCQF level 6)

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:
SQA Assessment Arrangements (www.sqa.org.uk/assessmentarrangements).

National Unit Specification: Statement of standards

Unit title: Machine Learning (SCQF level 6)

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.

Outcome 1

Demonstrate a conceptual understanding of machine learning and its applications.

Performance criteria

(a) Describe the concepts and terminology relating to machine learning.
(b) Describe real-world applications of machine learning.
(c) Describe challenges and limitations in the application of machine learning.
(d) Compare and contrast methods in common use in machine learning.

Outcome 2

Explain the role of algorithms in supervised machine learning.

Performance criteria

(a) Compare and contrast supervised and unsupervised machine learning.
(b) Explain the role that algorithms play in machine learning.
(c) Describe how a linear regression algorithm derives a predictive model for a variable.
(d) Describe how a decision tree models prediction for a categorical variable.

Outcome 3

Explain how the machine learning workflow contributes to model development.

Performance criteria

(a) Describe the elements of the machine learning workflow and their purpose.
(b) Explain the purpose of splitting data into training, validation and test sets.
(c) Describe common measures of model performance used in supervised learning.
(d) Explain the concepts of over-fitting and under-fitting, bias and variance.

National Unit Specification: Statement of standards (continued)

Unit title: Machine Learning (SCQF level 6)

Outcome 4

Apply the machine learning workflow to train and evaluate supervised learning models.

Performance criteria

(a) Describe the purpose of data preprocessing including data scaling and normalisation.
(b) Select and use appropriate software tools to preprocess data for machine learning.
(c) Select and use a linear regression machine learning algorithm to train a regression model on a dataset.
(d) Select and use a machine learning algorithm to train a binary classification model on a dataset.
(e) Evaluate the performance of a trained model.

Evidence requirements for this unit

Evidence is required to demonstrate that learners have achieved all outcomes and performance criteria.

Learner must provide **knowledge** and **product** evidence.

The knowledge evidence relates to all four 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). The amount of evidence may be the minimum required to infer competence across all outcomes.

The knowledge evidence may be sampled when testing is used. Testing must be carried out under supervised conditions, and it must be controlled in terms of location and time. Access to reference material is not permitted. The sampling frame, on all occasions, must include all four outcomes (but not every performance criterion within each outcome). The sampling frame must always include outcome 1, performance criterion (c).

National Unit Specification: Statement of standards (continued)

Unit title: Machine Learning (SCQF level 6)

The **product evidence** will relate to outcome 4. It will demonstrate that the learner has the competence and understanding to deal with datasets to prepare them for machine learning, and then select and apply a machine learning algorithm and interpret its results. The product evidence must satisfy the following criteria:

- Demonstrates use of appropriate analytic tools to preprocess data for machine learning.
- Produces a trained linear regression model using a machine learning algorithm on a dataset.
- Produces a trained binary classification model using a machine learning algorithm on a dataset.
- Evaluates the performance of a trained machine learning model.

This evidence may be produced over the life of the unit, under loosely controlled conditions (including access to reference materials). The datasets will be provided by the centre, chosen to match the performance criteria being evidenced.

The SCQF level of this unit (level 6) 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.

National Unit Support Notes

Unit title: Machine Learning (SCQF level 6)

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

The purpose of this unit is to develop understanding of the concepts that underpin machine learning and the methods used to implement it. It provides a practical introduction that will develop basic skills in training and evaluating common machine learning models. The machine learning workflow will give learners experience of data preprocessing, model training and evaluation of model fit.

There are three outcomes that build knowledge and understanding of the key elements of supervised machine learning, including approaches to optimisation and measuring model performance. The importance of good quality data is emphasised. The final outcome provides opportunity for the practical application of these concepts to real datasets, so that learners will experience most of the stages in the machine learning workflow.

No prior knowledge of machine learning is required, although learners should have familiarity with coding concepts and skills in numeracy. By the end of the course, learners will be able to independently implement machine learning workflows to solve problems with structured data.

For those learners undertaking this unit as part of the NPA in Data Science at SCQF level 6, this unit has strong links to the concepts that learners will encounter in Data Citizenship (SCQF level 6), and Data Science (SCQF level 6).

Learners will recognise the importance of the unit content to modern business, health and science where the use of big data and artificial intelligence methods are revolutionising operations and decision-making.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 6)

Individuals who successfully complete this unit will have an informed view of any claims made for predictions arising from a machine learning model. The knowledge and skills developed in this unit all have direct relevance to the practice of machine learning for supervised learning. Their application will be practised through practical problem solving and the use of computational methods.

Learners may progress to further study in computer science, data science, and artificial intelligence at SCQF level 7.

Knowledge	Skills
• Fundamentals of machine learning • Societal impacts and ethical considerations in machine learning applications • Importance of good quality data • The machine learning workflow • Data preprocessing and feature selection • Training, validation and test data • Algorithms, supervised and unsupervised learning • Comparison of statistical modelling and machine learning • Measuring trained model performance • Over- and under-fitting • Predictive modelling – continuous and categorical • Binary trees • Simple linear regression • Logistic regression • Bias-variance trade-off	• Select and use a software tool to preprocess data for machine learning • Fit a regression model to data • Fit a classification model to data (binary classification) • Evaluate trained model performance

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 6)

Guidance on approaches to delivery of this unit

You should have competence in applying the machine learning workflow to datasets with a view to prediction (linear regression) or binary classification (logistic regression). You should possess good data skills and have familiarity with the range of software environments and products that support machine learning. You should provide learners with access to machine learning software or online services. Your approach to delivery will be learner-centred, participative and practical. You should present and illustrate concepts through short case studies, using examples with a manageable number of features (explanatory variables). Where possible, you should encourage learners to practice data manipulation skills as the unit progresses.

You should use datasets that have well-understood characteristics to aid the learning of the methods and approaches and correspond well to the concepts being learned. There are many sources for datasets of this kind, and you should assemble and check these prior to the start of the course to ensure that they do not present problems that are beyond the scope of SCQF level 6. One useful source of datasets for this purpose can be found at Trello Data Sets.

In the case where learners have sufficient coding skills you might use coding environments such as Python or R Studio, along with appropriate libraries for machine learning and data analysis such as scikit-learn; matplotlib; caret; and ggplot2. You should also give learners the opportunity to hear from industry experts. For learners who do not have any coding skills, tools such as Orange or RapidMiner can be used.

The following distribution of time is suggested.

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

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.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 6)

Knowledge evidence for this unit could be assessed by methods such as:

1. A selected response test consisting of four options (one key) with a pass mark of 60%. Given that some performance criteria relate to explanations (rather than descriptions), there may need to be some scenario type questions to assess the learner's competency. The test could consist of a relatively high number of questions (30 or 40 for example), lasting an hour, which would span all of the outcomes and sample all of the knowledge statements (including at least one question for each statement).

Or

2. A constructed response test comprising a number of short answer questions, marked and assessed traditionally. For example, the test may comprise of 10 questions, requiring a response comprising no more than one or two paragraphs, selected across outcomes 1, 2 and 3, each worth five marks, with the learner responses marked out of 50 and a pass mark of 25. If this approach is taken, it is recommended that some (or all) of the questions 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 60 minutes.

Or

3. A report or presentation.

A more contemporary approach to assessment would involve the use of a blog or portfolio to record learning (and the associated activities) throughout the life of the unit. The blog would provide knowledge evidence (in the descriptions and explanations) and product evidence (using, for example, screenshots from package runs). The blog could be assessed using defined criteria to permit a correct judgement about the quality of the evidence. In this scenario, every performance must be evidenced; sampling would not be appropriate.

Where assessment evidence is generated out with class time the following strategies can be effective in authenticating a learner's work:

*   oral questioning.
*   write-ups under controlled conditions.
*   use of personal logs.
*   peer reports.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 6)

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 SQA e-Assessment. (www.sqa.org.uk/Guide to best practice.pdf).

Opportunities for developing Core and other essential skills

This unit provides opportunities to develop Core Skills.

Outcome 4 presents opportunities for learners to demonstrate the Core Skills of: Numeracy (handling data); Communication (presenting a report); Information and Communication Technology (ICT) (use of software to manipulate data and create reports) and Problem Solving (making decisions about model form and model adequacy, feature engineering).

This unit also develops computational thinking, such as:

- skills in abstraction (model building).
- decomposition (machine learning workflow).
- pattern recognition (model selection), and
- generalisation (using predictive data)

The skills developed are those in demand by employers seeking learners with good data skills. The broad understanding of the applications and limitations of machine learned modelling will serve to enhance citizenships skills (data citizenship).

History of changes to unit

Version	Description of change	Date

Unit template: February 2024

General information for learners

Unit title: Machine Learning (SCQF level 6)

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.

This is a non-specialist unit in which you will acquire a good understanding of machine learning concepts and its applications. You will learn about the principles that guide the selection of a model to train and apply a machine learning algorithm to train a model for prediction and a model for binary classification. To succeed in this unit, you should have a sound grasp of mathematical ideas and methods of carrying out computation.

You will acquire knowledge and skills in the following aspects of machine learning: basic concepts; impacts on society and ethical considerations. You will learn about the importance of good quality data, the danger of bias and the use of training, validation and test data in the machine learning workflow. You will understand the role of algorithms in supervised and unsupervised learning, and the difference between statistical modelling and machine learning. You will know how to measure model performance and understand the concepts of under- and over-fitting in model development.

This unit will develop your practical skills in using software to manipulate data and then train a model on it using a given algorithm. You will develop skills in measuring model performance and reporting on the outcomes.

The knowledge elements of this unit may be assessed by a knowledge test. This may be an online test, or a classroom test. The assessment of your skills in machine learning will comprise the analysis of given datasets. You will demonstrate skills in analysing data, training a model on data using an algorithm and reporting on the results from your analysis and model building.

The learning experiences in this unit will afford opportunity for you to develop the Core Skills of Numeracy (handling data); Communication (presenting a report); Information and Communication Technology (ICT) (use of software to manipulate data and create reports) and Problem Solving (making decisions about model form and model adequacy, feature engineering).

You will also develop computational thinking, such as:

- skills in abstraction (model building).
- decomposition (machine learning workflow).
- pattern recognition (model selection), and
- generalisation (using predictive data).

General information for learners (continued)

Unit title: Machine Learning (SCQF level 6)

The broad understanding of the applications and limitations of machine learned modelling will serve to enhance your citizenships skills (data citizenship).

You may progress to units in Data Science at higher levels, such as: Machine Learning (SCQF level 7); Artificial Intelligence (SCQF level 7) and Big Data (SCQF level 7).