National Unit Specification

General information

Unit title: Machine Learning (SCQF level 5)

Unit code: J891 45

Superclass: CB

Publication date: June 2024

Source: Scottish Qualifications Authority

Version: 01

Unit purpose

The purpose of this unit is to provide a straightforward introduction to Machine Learning and its applications.

This is a non-specialist unit, intended for learners with a vocational interest in STEM. It is particularly suitable for learners who have experience of other computer science or data science topics and wish to broaden their understanding of the discipline. It will also serve learners who wish to understand the basic principles of Machine Learning.

No previous knowledge or experience of Machine Learning is required. However, it is recommended that learners have computational and numerical skills before undertaking this unit.

Learners will be introduced to the purpose of Machine Learning and the role of algorithms in learning from data. They will understand the importance of good quality data and the role of feature selection in the Machine Learning process. They will distinguish between supervised and unsupervised learning and identify the role of training, validation and test datasets in the machine learning process. Learners will explain the impacts of the use of Machine Learning in decision making and the ethical issues this can raise.

Learners may wish to progress to further study in related topics, such as Data Science: Machine Learning SCQF level 6 (J891 46).

National Unit Specification: General information (continued)

Unit title: Machine Learning (SCQF level 5)

Outcomes

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

1. Describe the basic concepts of machine learning and the factors that enable it.
2. Describe applications of machine learning and the benefits and challenges that arise from its use
3. Describe the machine learning workflow.
4. Demonstrate understanding of binary classification.

Credit points and level

1 National Unit credit(s) at Scottish Credit and Qualifications Framework (SCQF) level 5: (6 SCQF credit points at SCQF level 5).

Recommended entry to the unit

Entry is at the discretion of the centre. No previous knowledge or experience of Machine Learning is required. Basic number skills are assumed.

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.

The target cohort is learners who are choosing an optional unit as part of the NPA in Data Science at SCQF level 5.

The unit may also be of interest to learners wishing to further their knowledge and understanding in computer science or data science.

National Unit Specification: General information (continued)

Unit title: Machine Learning (SCQF level 5)

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 5)

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

Describe the basic concepts of machine learning and the factors that enable it.

Performance criteria

(a) Describe the basic concepts of machine learning.
(b) Describe the factors that have enabled the growth in machine learning.
(c) Describe how machine learning relates to deep learning and natural language processing.
(d) Describe the differences between coding a statistical model and training a model through machine learning.
(e) Describe the differences between organisational use of machine learning models and large language models (LLMs).

Outcome 2

Describe applications of machine learning and the benefits and challenges that arise from its use.

Performance criteria

(a) Describe common applications of machine learning in various sectors.
(b) Describe potential benefits from the use of machine learning in various contexts, including large language models.
(c) Describe potential disadvantages in the application of machine learning.
(d) Describe ethical issues that can arise through the use of machine learning models.
(e) Describe sustainability implications of machine learning.

Outcome 3

Describe the machine learning workflow.

Performance criteria

(a) Describe the initial stages of data preparation for a machine learning method.
(b) Describe common methods for transforming data for use in a machine learning context.
(c) Describe the purpose of training, validation and test data splits.
(d) Describe the role of algorithms in training a machine learning model.
(e) Describe the role of model evaluation in relation to model deployment.

National Unit Specification: Statement of standards (continued)

Unit title: Machine Learning (SCQF level 5)

Outcome 4

Demonstrate understanding of binary classification.

Performance criteria

(a) Explain the basic principle of a binary classification machine learning algorithm.
(b) For given data, apply an algorithm to train a binary classification model, following the steps of the machine learning workflow.
(c) Evaluate a binary classification model trained on given data.

Evidence requirements for this unit

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

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

The knowledge evidence relates to all 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 performance criteria.

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 a minimum of two performance criteria from each outcome.

The product evidence will relate to performance criteria in outcome 4. It will demonstrate that the learner can follow the steps of the machine learning workflow, from data acquisition to model fitting and evaluation. They should demonstrate use of appropriate software tools in regular use for machine learning. In particular, the evidence must demonstrate that the learner can:

- preprocess a dataset for a binary classification application of machine learning.
- transform data to enable processing in a machine learning context.
- create a data split for training and validation.
- use an algorithm to train a binary classification model.
- evaluate the performance of the trained model (accuracy and precision).

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 datasets will be provided by the centre, chosen to match the performance criteria being evidenced.

National Unit Specification: Statement of standards (continued)

Unit title: Machine Learning (SCQF level 5)

The SCQF level of this unit (level 5) 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.

National Unit Support Notes

Unit title: Machine Learning (SCQF level 5)

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

This unit provides an introduction to the basic principles that underpin machine learning, along with an appreciation of its power in supporting decision making in areas such as business, health and science. Learners will also be made aware of the ethical challenges that arise from the use of machine learning. They will demonstrate competence in the application of machine learning to a binary classification problem.

This unit may be delivered as a stand-alone unit or in combination with other units as part of a group of units making up an award, such as NPA in Data Science at SCQF level 5. This unit is suitable for learners who have no previous experience of machine learning or who wish to enhance their knowledge and further develop their skills in the area of data science.

The unit will cover the following knowledge and skills.

Knowledge	Skills
• Concept of Machine Learning (ML) and its history • ML methods • ML versus traditional programming • Supervised, unsupervised learning • Applications of ML • Supervised learning — linear regression • Unsupervised learning — clustering • Neural networks and deep learning • Ethical challenges of ML (privacy and transparency)	• Identify type of machine learning from given scenario • Pre-process data, including transformation • Create a data split for training and validation • Train a binary classification model from given data • Evaluate the performance of a trained binary classification model

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 5)

Knowledge	Skills
• ML workflow (raw data to deployment) • Datasets for training ML models • Training, validation and test data split • Data preprocessing and transformation • Algorithms and model training • Binary classification model and algorithms • Model accuracy metrics (accuracy, precision, F1) • How LLMs learn from text data • LLMs and text generation	

Guidance on approaches to delivery of this unit

This unit is an optional unit for the National Progression Award in Data Science at SCQF level 5. As such, it may be delivered alongside the mandatory units: Data Citizenship (SCQF level 5) and Data Science (SCQF level 5). In this circumstance, teaching, learning and assessment may be integrated across the units.

In outcome 1 you should set machine learning in the context of other artificial intelligence methods, and its relationship to these methods explained. This will provide opportunity for you to explain the characteristics of machine learning and its relationship to big data and contrast it with traditional approaches to coding solutions to problems. You should provide readily available practical examples of machine learning use, including image processing and text processing as demonstrated in large language models.

The history of machine learning development can be briefly covered. You should keep the treatment of large language models as light as possible, such as a high-level overview explaining how they work (large neural nets trained on text data) and their capabilities (generating text, answering questions, summarising). The goal should be conveying the transformative potential these technologies have rather than the technical details.

You should use outcome 2 to enable learners to understand that while machine learning enables many practical applications that provide benefits in business, science and health there are also concerns related to the ethics of using data in this way.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 5)

You should provide examples where machine learning helps decision making. These should include image recognition (security surveillance, medical imaging), business intelligence (sales and marketing), healthcare diagnosis and personalised treatment plans) and speech recognition (dictation, virtual assistants). But these also carry risks. You should explain that machine learning algorithms used to analyse user behaviour and preferences on social media platforms like YouTube, Instagram, and Facebook and may reinforce prejudices or misconceptions. Face recognition may incorrectly identify individuals as suspects, and medical diagnosis could be incorrect leading to the wrong treatment being applied.

You should touch more lightly on aspects of machine learning that have ethical implications, such as concerns around personal data privacy and bias. While machine learning models lead to more efficient processes in business and engineering, you should point out that it is an energy intense process leading to issue of sustainability.

In delivering outcome 3 you should briefly cover each of the stages of the machine learning workflow from data ingestion to model deployment. Provide opportunity for learners to explore sample datasets that illustrate typical problems with incomplete or inaccurate data and the methods to improve data quality. You should introduce the purpose of training, validation and test datasets and provide illustrations of feature selection through data analytics. You should spend time explaining the role of algorithms in machine learning, along with illustrations of their operation in basic terms.

In delivering outcome 4, you can briefly cover the main types of classification algorithm before moving on to explain one in sufficient depth along with worked-through practical illustrations. You should introduce the confusion matrix along with the various measures of performance that it gives rise to. You should illustrate the concept of overfitting in a binary classification model. You should provide learners with the software tools they need to preprocess data, create dataset splits and train a binary classifier model for a dataset.

You could use free tools such as Orange or RapidMiner that allow hands-on experimentation and model building through graphical interfaces, requiring minimal coding. Features like visualisation, sample datasets, and predefined algorithms help to make machine learning more accessible for an introductory course. The Machine Learning for Kids website shows how learners can use the TeachableMachine tool and Scratch block programming to explore machine learning.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 5)

The following distribution of time is suggested.

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

Guidance on approaches to assessment of this unit

The following are suggestions only. There may be other methods that would be more suitable to learners and the type of learner assessment activities will vary depending on the resources available.

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.

Most of the assessment evidence in this unit relates to knowledge and understanding. The traditional approach to assessment would comprise a test, taken towards the end of the unit. The test could consist of a number of selected response questions, chosen from all of the outcomes and performance criteria. Not every performance criterion should be tested but every outcome could be tested by a number of questions.

It is suggested that some elements of knowledge lend themselves to a form of selected response (multiple choice) questions. Where explanations are required, the form of the question could be constructed response, scenario-based where appropriate. The whole test could be timed and carried out under controlled conditions, without access to reference material. Where re-assessment is required it should contain a significantly different sample selected from the range of mandatory content.

In the case of the product evidence, learners should be supplied with datasets that have been pre-prepared by the centre to allow the learner to demonstrate that they can select and use appropriate software tools to achieve the set tasks and report on the outcomes. An assessor observation checklist could be used to record that the assessment tasks have been undertaken successfully by the learner.

National Unit Support Notes (continued)

Unit title: Machine Learning (SCQF level 5)

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

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 adequacy).

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 5)

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 unit provides a straightforward introduction to machine learning and its applications. It assumes no prior knowledge of machine learning but requires that you have a good foundational knowledge in mathematical topics, including basic statistics, and can understand how computers process information.

The unit will explain the broad purpose of machine learning and its relationship to other artificial intelligence methods, and introduce you to typical applications in business, science and health. You will appreciate that machine learning brings benefits but also risks to the individual. The unit will cover the role of good quality data in the machine learning workflow, how algorithms function to train a model and the uses to which those models can be put.

In outcome 1 you will learn the basic concepts of machine learning including large language models.

In outcome 2 you will learn about the machine learning workflow from raw data to a deployed application.

In outcome 3 you will learn about binary classification and train a binary classifier on a dataset.

In outcome 4 you will learn about the range of applications of machine learning and the ethical and other challenges that arise from their use.

The unit covers the following knowledge and skills.

Knowledge	Skills
• Concept of Machine Learning (ML) • ML methods • ML versus traditional programming • Supervised, unsupervised learning • Applications of ML • Supervised learning — linear regression • Unsupervised learning — clustering • Neural networks and deep learning	• Identify type of machine learning from given scenario • Train a binary classification model from given data • Evaluate the performance of a binary classification model.

General information for learners (continued)

Unit title: Machine Learning (SCQF level 5)

Knowledge	Skills
• Ethical challenges of ML (privacy and transparency) • ML workflow (raw data to deployment) • Datasets for training ML models • Training, validation and test data split • Data preprocessing and feature selection • Algorithms and model training • Binary classification model and algorithms • Model accuracy metrics • How LLMs learn from text data • LLMs and text generation	

You can be assessed in a variety of ways. The knowledge and understanding components of the unit may be assessed through a knowledge test as determined by your centre, in controlled conditions. Typically, this will involve a blend of multiple-choice questions, short response questions and scenario-based questions.

The practical elements of the unit will be assessed by applying machine learning to a binary classifier problem.

The unit will provide an opportunity for you to progress in the Core Skills of Numeracy, Problem Solving and Information and Communication Technology (ICT). You will also progress in computational thinking through exposure to algorithms and their operation. Your understanding of the implications of machine learning for society and the individual will contribute to your data citizenship competences.

On completion of this unit, you may progress to Data Science: Machine Learning SCQF level 6. You will also have the knowledge and skills to undertake a project in Machine Learning in the unit J2GT 45 Data Science Project at SCQF level 5.