

Overview

This standard identifies the competences you need to assist in the development of machine learning algorithms and their implementation, in accordance with approved procedures.

You will be required to assist in the development of machine learning algorithms, which encompass machine learning workflow; supervised and unsupervised learning; creation of training and test datasets; fitting classifier, regression and clustering models and interpreting the results.

Machine learning algorithms are used in a wide variety of applications where it is difficult or infeasible to develop a conventional algorithm to perform the task. This will involve the practical use of software tools for machine learning algorithm development. You will be able to assess the performance of a developed model and identify the role of training and test datasets in this process.

Your responsibilities will require you to comply with organisational policies and procedures. You will be expected to work to instructions, alone or in conjunction with others, taking personal responsibility for your own actions.

Your underpinning knowledge will be sufficient to provide a sound basis for your work and will enable you to apply the required procedures for the development and testing of machine learning algorithms. You will recognise the importance of good quality data and the role of feature selection in the learning process.

This activity can be increasingly found in any sector or organisation and in particular those associated with implementing automated reasoning systems that can learn to respond based on training datasets provided. It is likely to be undertaken by people working as Junior Machine Learning Specialists or Junior Machine Learning Engineers.

Performance criteria

You must be able to:

1. perform data extraction, preparation and transformation in order to produce required datasets
2. assist with data cleaning of noisy, incomplete data or data with established data quality issues using approved tools and techniques
3. assist in selecting and applying statistical tools to generate descriptive statistics from different datasets for the purpose of feature selection
4. assist with creating analytical models using approved modelling techniques to model structured data
5. assist in applying best-practice model fit testing and validation techniques to assess model performance
6. select a classifier algorithm, and use industry standard software tools to load a dataset and produce a machine learning classifier model
7. apply feature selection and linear methods in order to perform variable reduction to improve the performance of models
8. assist in evaluating the measures of model fit for a classifier model
9. assist in identifying the features of a classifier model
10. document the modelling process in line with organisational standards
11. assist in producing documentation in order to secure implementation sign-off
12. produce visualisations, charts and graphs to communicate data and model process in required timescales

Knowledge and understanding

You need to know and understand:

1. the purpose, key features and applications of machine learning
2. the role that algorithms play in machine learning
3. the need to validate machine learning models and how to carry this out
4. the basic concepts of variable creation and reduction in data analysis
5. how variables and features impact model performance in testing and validating analytical models
6. the potential data quality issues that can arise, including missing values, duplicate data, incorrect data and how to deal with these
7. the implications of data quality for analysis model performance
8. the importance of feature selection in effective machine learning
9. the industry standard approaches used for data cleansing and how to apply them
10. the difference between supervised, unsupervised and reinforcement learning
11. the purpose of training and testing data sets in developing and evaluating a machine learning model
12. the feature identification stage used in model development and how to perform this
13. the industry standard statistical methods and best-practice modelling techniques (including classifier, regression and clustering models) used to develop machine learning solutions
14. the main types of algorithm used to develop machine learning solutions, including decision trees, nearest neighbour and linear classifier
15. the industry standard tools used for implementing algorithms and developing machine learning models
16. the characteristics of under-fitting and over-fitting in a classifier model
17. the different approaches to model improvement in a classifier problem
18. the measures of performance that can be used in model development, covering classification, prediction and clustering

TECIS805301

Assist in developing and validating machine learning solutions



Developed by	ODAG
Version Number	1
Date Approved	April 2020
Indicative Review Date	March 2023
Validity	Current
Status	Original
Originating Organisation	ODAG Consultants Ltd
Original URN	TECIS805301
Relevant Occupations	Information and Communication Technology Professionals; Software Development
Suite	IT(Data Science)
Keywords	Machine learning, algorithms, artificial intelligence
