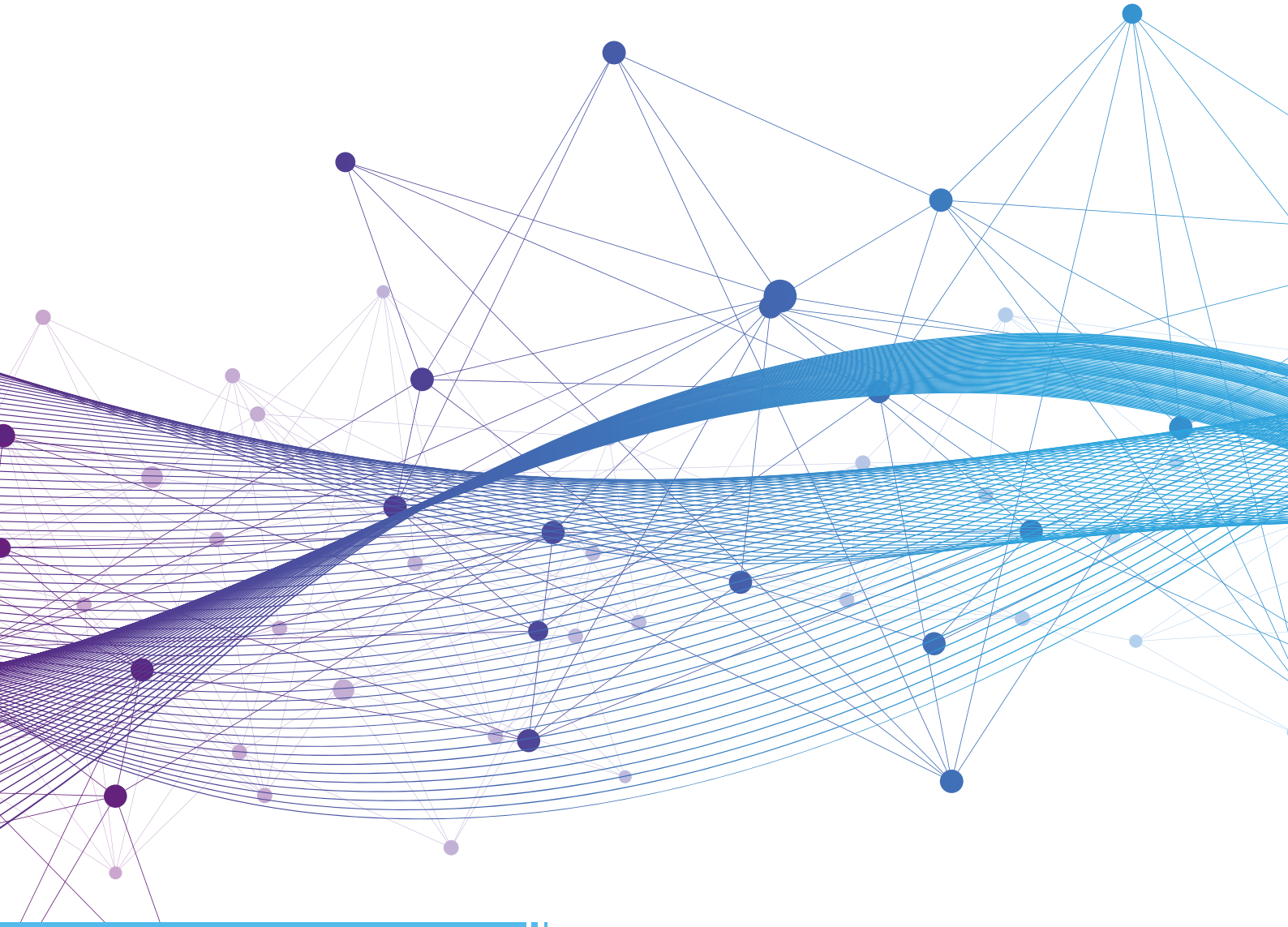
DELTA
ANALYTICS TRAINING

Level 3 International Diploma (Data Science)

Syllabus 2024



Change *your* future
www.delta-training.co.uk

AIM Qualifications and Assessment Group

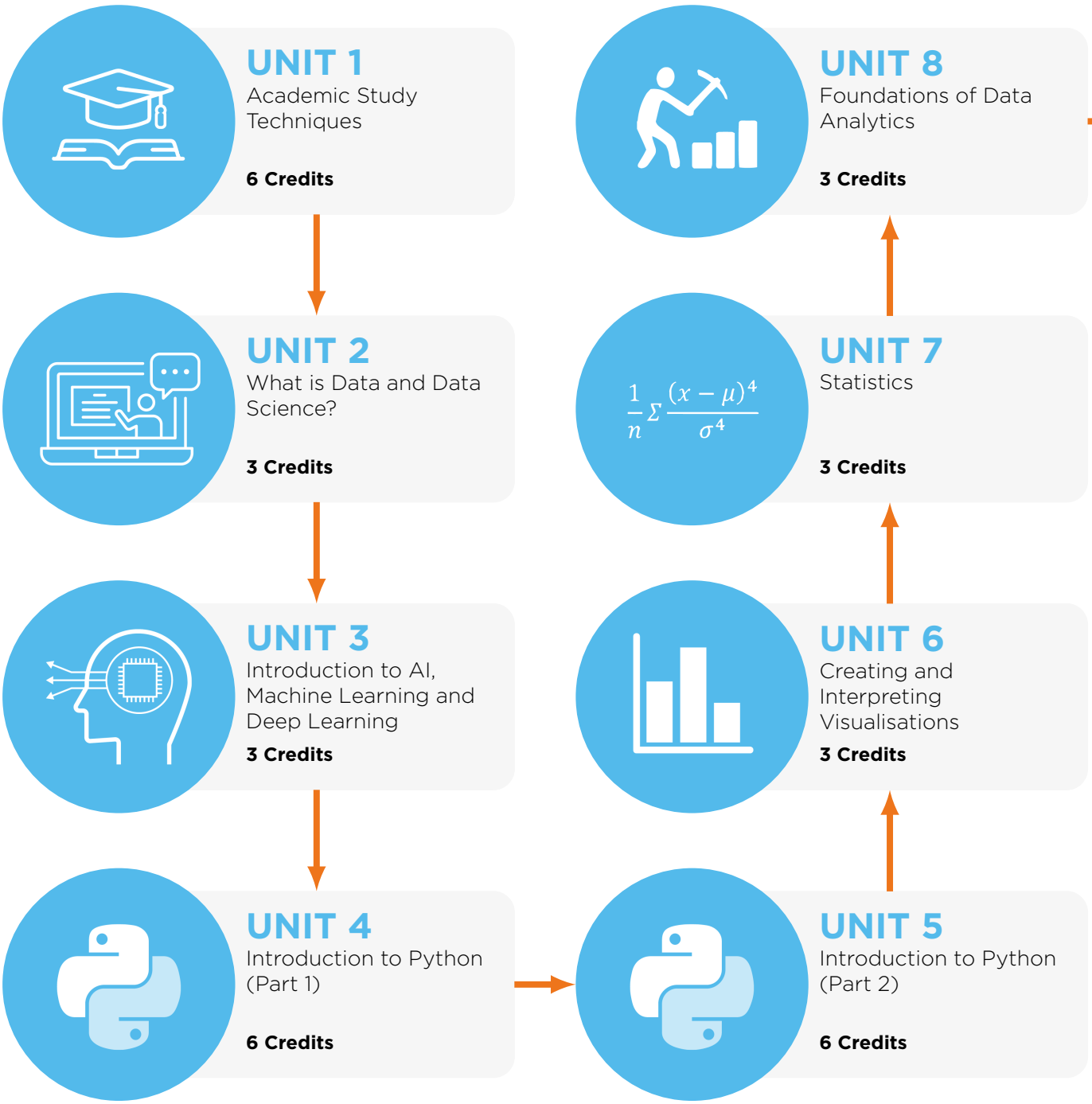
Introduction

This document details the 2024 syllabus (i.e. learning outcomes and assessment criteria) for the **Level 3 International Diploma (Data Science)**. This document applies to **all overseas students** studying the diploma.

For detail on how to register for the diploma, our entry requirements, details of our distance learning delivery method, and the tutor and wellbeing support we offer, please read the **Diploma Information 2024** booklet associated with this version of the diploma.

The learning pathway

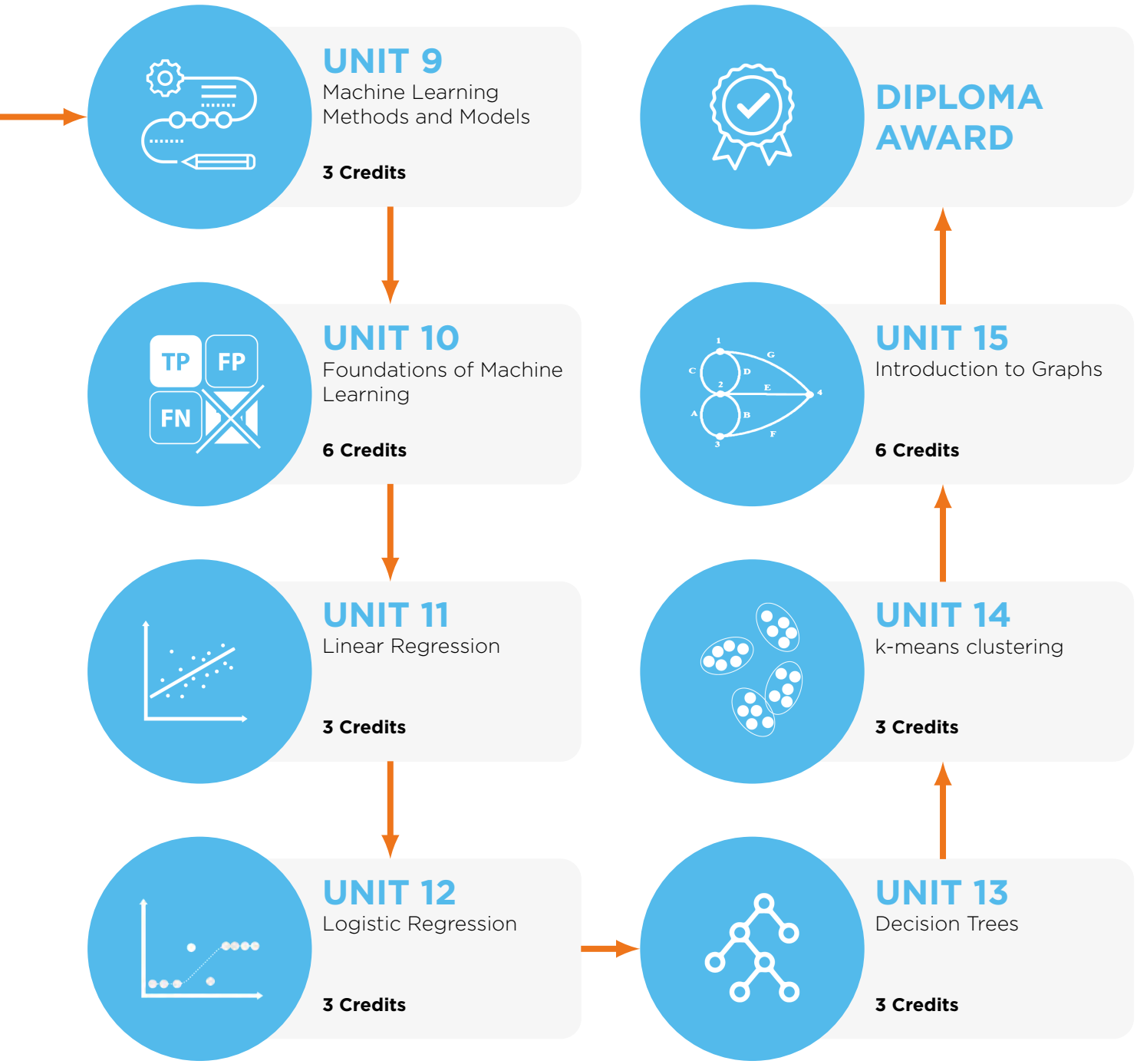
Over 15 mandatory units, the diploma provides a contemporary introduction to data science, artificial intelligence, and machine learning, taking students on a journey from the birth of AI in the 1960s, right through to the big data era of the early 2000s.



On the journey, students will learn about the current applications of AI and machine learning and the various challenges associated with them.

In addition to learning about standard linear regression, logistic regression, decision trees and k-means clustering, the diploma introduces graph data science, an exciting and emerging area of data science.

The diploma also introduces students to the data analytical landscape and associated analytical tools, teaching introductory Python so that students can analyse, explore, and visualise data, as well as train, test and apply basic data science models.



Unit 1 – Academic Study Techniques

6 Credits | 60 hours

Unit aims

This unit prepares students for academic learning, by providing practical advice to help learn, build and extend skills in all areas of study, including studying, scientific writing, note-taking and referencing.

Learning Outcomes and Assessment Criteria

1. Understand personal learning strengths and weaknesses.

1.1 Evaluate personal learning strengths and weaknesses and outline a strategy for improvement.

2. Understand how to plan and organise study.

2.1 Discuss factors affecting the planning and organisation of study in relation to personal circumstance.
2.2 Develop own personal time management schedule.
2.3 Evaluate own personal time management schedule.

3. Be able to assess the reliability of sources of information.

3.1 Assess three secondary sources for their:

- Authenticity
- Currency
- Credibility
- Relevance
- Accuracy
- Audience/purpose

4. Understand how to use reading strategies.

4.1 Use and evaluate three reading strategies.

5. Understand note-taking methods.

5.1 Evaluate three note-taking methods.
5.2 Produce effective notes from three different sources.

6. Understand how to plan written work.

6.1 Use and review planning techniques in written work.

7. Understand how sources are referenced.

7.1 Explain a variety of ways in which sources are referenced.
7.2 Use references correctly following a recognised system.
7.3 Explain why and how plagiarism must be avoided

8. Be able to write with relevance in responding to an academic task.

8.1 Demonstrate the ability to select and incorporate relevant information in a written (academic) task.

Unit 2 – What is Data and Data Science?

3 Credits

30 hours

Unit aims

This unit introduces students to the field of data science from its beginnings in the 1950s as an alternative phrase for "computer science" to its emergence as a profession in the mid 2010s. The unit also introduces students to "big data" and how it has evolved over the last decade or so. Finally, the unit introduces a number of the common tools used by data scientists.

Learning Outcomes and Assessment Criteria

1. Understand the core issues of data science.

1.1 Explain what is meant by the terms **data science** and **data scientist**.
1.2 Explain how **data science** is related to other academic fields.
1.3 Discuss three applications of **data science**.

2. Understand the core issues of data and big data.

2.1 Explain what is meant by the terms **data** and **big data**.
2.2 Analyse the challenges and criticisms of **big data**.
2.3 Analyse two successes and two failures of **big data**.

3. Understand the data science ecosystem.

3.1 Compare and contrast common tools and software used by **data scientists**.
3.2 Compare and contrast tools used to process and analyse **big data**.

Unit 3 – Introduction to AI, Machine Learning and Deep Learning

3 Credits | 30 hours

Unit aims

This unit introduces students to artificial intelligence and machine learning from their development in the 1950s, to their current applications and their various challenges. The unit also introduces deep learning models, examining their architecture and reviewing their uses and limitations.

Learning Outcomes and Assessment Criteria

1. Understand the core issues of artificial intelligence.

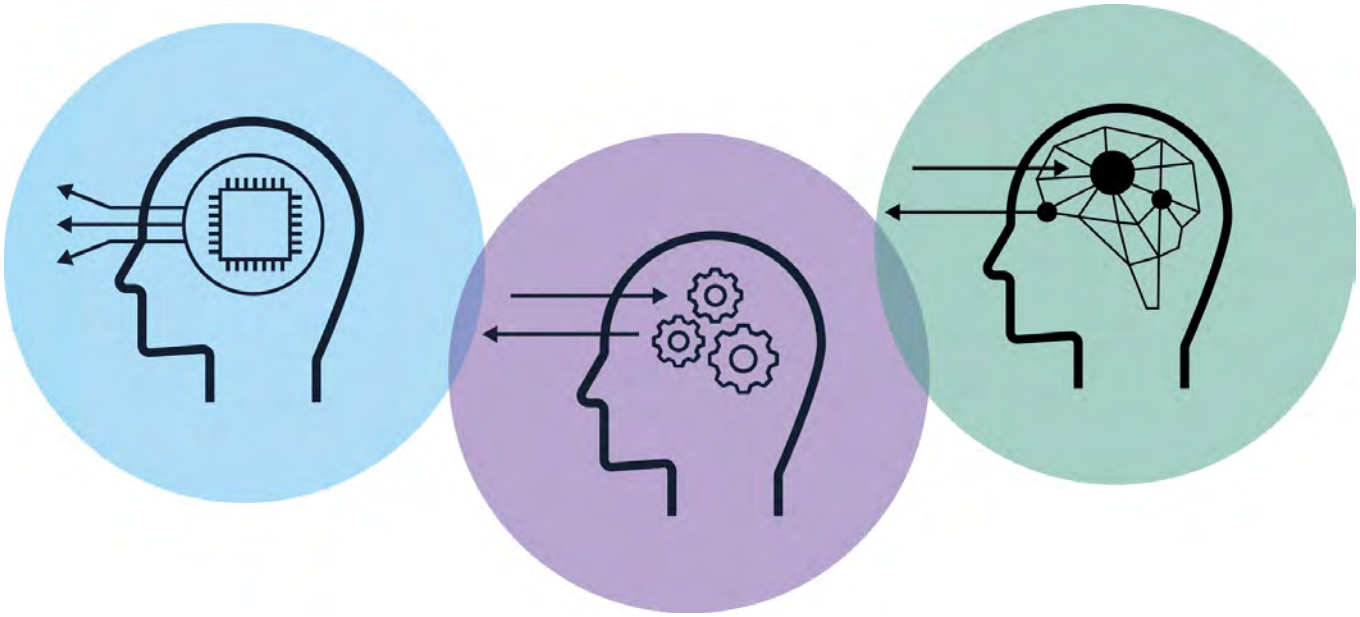
1.1 Define what is meant by the term "artificial intelligence".
1.2 Explain the difference between the terms "artificial narrow intelligence", "artificial general intelligence" and "artificial super intelligence".
1.3 Analyse the challenges in achieving artificial intelligence.
1.4 Analyse the successes and failures of artificial intelligence.

2. Understand the core issues of machine learning.

2.1 Define what is meant by the term "machine learning".
2.2 Explain the main types of machine learning: "supervised", "unsupervised" and "reinforcement learning".
2.3 Analyse the uses and limitations of machine learning.
2.4 Explain the difference between artificial intelligence and machine learning.

3. Understand the core issues of deep learning.

3.1 Define what is meant by the term "deep learning".
3.2 Explain deep learning architecture.
3.3 Analyse the uses and limitations of deep learning.
3.4 Analyse three current areas of research in deep learning.



Unit 4 – Introduction to Python (Part 1)

6 Credits

60 hours

Unit aims

This unit introduces students to Python programming for data science. The unit assumes no prior knowledge of coding or of Python and so starts by explaining the basics of Python, its design philosophy, syntax, naming conventions and coding standards. The unit then introduces some basic Python data types and explains how these data types can be created, changed, manipulated, and calculated using standard mathematical functions, logical operators, and Python's built-in methods.

Learning Outcomes and Assessment Criteria

1. Understand the design philosophy and core features of Python.

1.1 Explain what is meant by Python being a "*high-level, interpreted, dynamically-typed, general-purpose language*."

1.2 Compare and contrast Python to two programming languages such as C++ and R.

1.3 Explain Python's syntax, indentation, naming conventions and coding standards.

2. Understand the Python ecosystem.

2.1 Compare and contrast three of each of the following:

- Common text editors.
- IDEs.
- Development tools.

2.2 Discuss three Python key libraries and frameworks.

3. Understand Python's basic data types.

3.1 Explain the basic Python data types: strings, integers, floats, complex numbers, and Booleans.

3.2 Use arithmetical operators and standard mathematical functions correctly to perform basic calculations.

3.3 Use the logical, bitwise and identity operators correctly to perform logical operations.

3.4 Explain the order of operator precedence.

3.5 Use string methods and functions correctly to create new strings or to retrieve values and properties.

3.6 Obtain string elements correctly by indexing and slicing.

Unit 5 – Introduction to Python (Part 2)

6 Credits

60 hours

Unit aims

This unit builds on the previous unit by introducing more complex data structures critical to many data science tasks, such as “lists”, “tuples”, “sets”, and “dictionaries”. The unit also explains how to use control and flow statements such as branching and looping, as well as the basics of writing user-defined Python functions – all the ingredients needed to code data science models successfully.

Learning Outcomes and Assessment Criteria

1. Be able to create and manipulate lists.

1.1 Explain what a Python list is.
1.2 Create three lists of different data types.
1.3 Use list methods: append, clear, insert, extend, sort, reverse, remove, pop.
1.4 Obtain list elements by indexing and slicing.
1.5 Order a list using sort and sorted function.

2. Be able to create and manipulate tuples.

2.1 Explain what a Python tuple is.
2.2 Explain the difference between a list and a tuple.
2.3 Create tuples of five data types.
2.4 Use tuple methods: count, index.
2.5 Obtain tuple elements by indexing, slicing and tuple unpacking.

3. Be able to create and manipulate sets.

3.1 Explain what a Python set is.
3.2 Create sets of five data types.
3.3 Use four set operations and methods.

4. Be able to create and manipulate dictionaries.

4.1 Explain what a Python dictionary is.
4.2 Create dictionaries of five data types.
4.3 Retrieve keys and values from a dictionary.
4.4 Update existing values, add, and delete key-value pairs.

5. Be able to write and use basic Python functions.

5.1 Create def defined functions, passing parameters, and returning values.
5.2 Create anonymous or Lambda functions.
5.3 Explain the difference between keyword, positional and optional parameters.
5.4 Explain the differences between def defined functions and lambda functions.
5.5 Write docstrings in functions.
5.6 Write type hints to indicate the type of a parameter.

Unit 6 – Creating and Interpreting Visualisations

3 Credits

30 hours

Unit aims

This unit introduces students to basic charts and visualisations, and how to create and interpret them. The unit starts by explaining why visualisations are critical when understanding data and what makes a good and a poor visualisation. The unit also focuses on some of the Python libraries used to produce charts and visualisations.

Learning Outcomes and Assessment Criteria

1. Understand the importance of visualising data.

1.1 Explain the role and importance of visualising data before conducting data analysis.

1.2 Explain why poorly created visualisations can be misleading.

1.3 Explain good practices when creating plots and charts.

2. Understand basic plots and charts

2.1 Explain the advantages and disadvantages of three chart types from:

- Scatter plots.
- Line charts.
- Pie charts.
- Bar and column charts.
- Histogram and density curves.
- Box and whisker plots.

2.2 Explain which chart types should be used for different types of data.

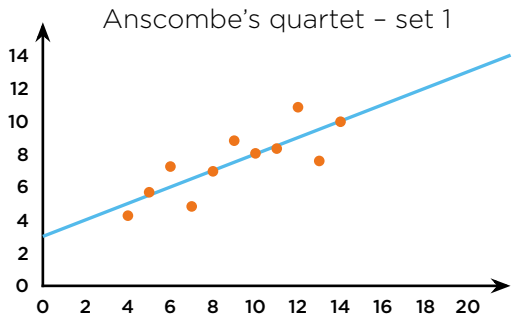
3. Understand features of Python Libraries.

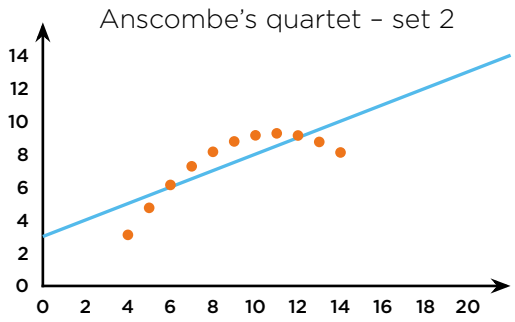
3.1 Analyse the use of Python libraries for constructing charts and visualisations.

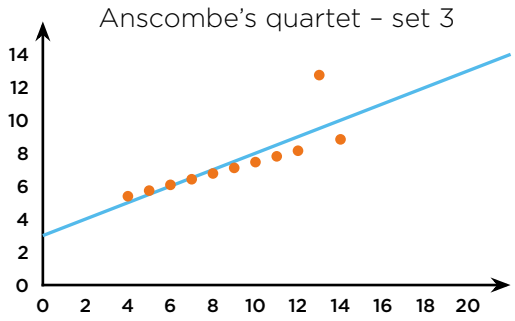
4. Be able to create and interpret plots and charts

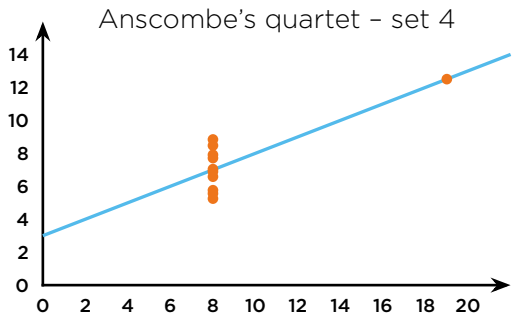
4.1 Write Python code to construct, format and display three basic charts and plots.

4.2 Interpret the charts produced from 4.1









Unit 7 – Statistics

3 Credits | 30 hours

Unit aims

This unit aims to provide students with an introduction to descriptive statistics and methods which are key for data analysis and data science. This unit introduces different types of data and descriptive statistics including measures of centre, spread, symmetry and measures of joint variability. The unit also explains which descriptive statistics can be calculated for the data measured on different scales.

Learning Outcomes and Assessment Criteria

1. Be able to describe a data set.

1.1 Calculate and interpret measures of location (mean, median, mode) for a variety of data sets.
1.2 Calculate and interpret measures of spread (standard deviation, range and IQR) for a variety of data sets.
1.3 Plot and use box plots to describe (e.g. skewness) and compare data sets.
1.4 Interpret and discuss the attributes of a data set as given by graphs, charts and diagrams (e.g. histograms, line graphs).

2. Analyse bivariate data.

2.1 Interpret scatter graphs arising from primary or secondary data.
2.2 Calculate and interpret the product moment correlation coefficient.
2.3 Calculate least squares regression lines (one explanatory variable), fitting this line to scatter diagrams and using the model.
2.4 Identify outliers and understand the impact of these on fitted regression lines.

3. Know about the binomial distribution.

3.1 Decide when and if a binomial model is appropriate.
3.2 Calculate probabilities using the binomial distribution formula.
3.3 Use formula to find the mean and standard deviation of binomial distribution and interpret in context.

4. Know about the normal distribution.

4.1 Decide when and if a normal model is appropriate.
4.2 Calculate probabilities using statistical tables.

Unit 8 – Foundations of Data Analytics

3 Credits

30 hours

Unit aims

This unit serves as the introduction to the core concepts of data analytics, including the fundamental tasks and processes in the data discovery process such as data cleaning, methods for dealing with data quality and basic methods for standardising data. The unit also introduces students to some of the different data file formats and databases commonly used by data scientists.

Learning Outcomes and Assessment Criteria

1. Understand the key processes and types of data analytics.

1.1 Explain the "Knowledge Discovery from Data" process.
1.2 Explain the different types of data analytics: **descriptive**, "predictive", and "prescriptive".
1.3 Explain the differences between the roles: "data engineer", "data analyst", "data scientist" and "business intelligence analyst".

2. Understand the data analytics ecosystem.

2.1 Compare and contrast different types of data format: CSV, JSON, Excel, text, audio, and images.
2.2 Explain the difference between relational and non-relational databases.
2.3 Compare and contrast three common software tools used for data analytics.

3. Understand the issues and methods for dealing with data quality issues.

3.1 Explain the strategies for identifying and dealing with:
- Missing data.
- Duplicate data.
- Inconsistent data.
- Outliers.

3.2 Compare and contrast "mean", "median" and "mode" strategies for data imputation.

4. Understand the issues and methods of basic data transformations.

4.1 Explain the purpose of three data transformation strategies: smoothing, feature engineering; aggregation; normalization; discretization.
4.2 Explain the difference between min-max and z-score normalization.
4.3 Explain how binning can be used to smooth data and to discretize data.

Unit 9 – Machine Learning Methods and Models

3 Credits

30 hours

Unit aims

This unit provides a high-level overview of the three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. The unit discusses the use-cases and real-world problems the various methods can be applied to, summarises the key-features of the different methods, as well as the challenges of each method.

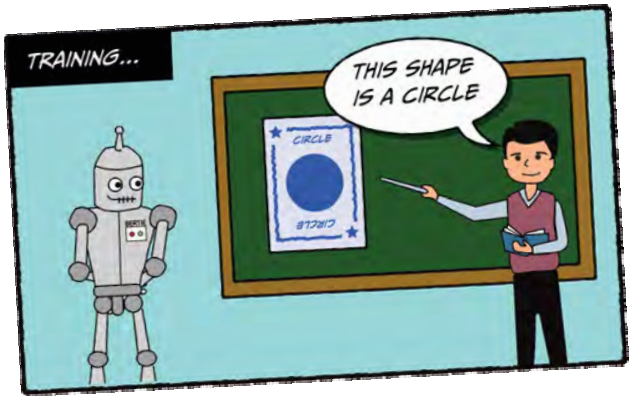
Learning Outcomes and Assessment Criteria

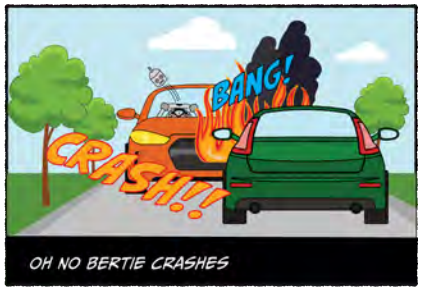
1. Understand the core concepts of basic supervised machine learning models.

1.1 Explain the key features and objectives of:
- Linear regression.
- Logistic regression.
- Decision trees and random forests.
- Support Vector Machines.
- K-Nearest Neighbour.

1.2 Discuss the challenges of supervised models.

1.3 Explain three types of use-cases to which supervised ML models can be applied.





2. Understand the core concepts of basic unsupervised machine learning models.

2.1 Explain the key features and objectives of:
- Clustering.
- Association rules.
- Dimensionality reduction.

2.2 Discuss the challenges of unsupervised models.

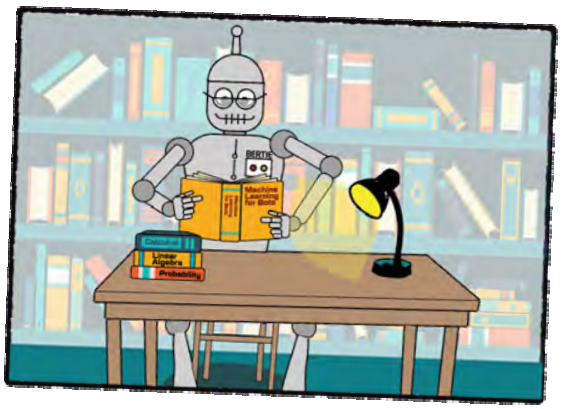
2.3 Explain the types of use-cases to which unsupervised ML models can be applied.

3. Understand the core concepts of reinforcement learning.

3.1 Explain the key features and objective of reinforcement learning.

3.2 Analyse the challenges of reinforcement learning.

3.3 Explain a use-case to which reinforcement learning can be applied.



Unit 10 – Foundations of Machine Learning

6 Credits

60 hours

Unit aims

This unit introduces the many steps and processes involved when building and evaluating machine learning models, including how to prepare data for machine learning algorithms, how to split data into training, test, and validation datasets, how to assess the accuracy of machine learning models, and how to identify and correct class imbalance.

Learning Outcomes and Assessment Criteria

1. Understand the machine learning process.

1.1 Analyse the components of the machine learning process: data collection; data preparation; selecting the machine learning algorithm; training and testing models; parameter tuning; deploying a model.

1.2 Explain the difficulties and solutions for each component of the machine learning task.

2. Understand the data preparation process for machine learning models.

2.1 Compare and contrast the data requirements for different machine learning models.

2.2 Explain how to convert categorical data to numerical values.

2.3 Explain why "class imbalance" can be dangerous for models.

2.4 Compare and contrast the following strategies for balancing classes:
- Over sample the minority class.
- Under sample the majority class.

2.5 Explain the purpose of splitting data into training, test and validation subsets.

3. Understand how to evaluate machine learning models.

3.1 Explain what is meant by a "confusion matrix".

3.2 Define the classification metrics: "Precision", "Accuracy", "Recall", "Support" and "F1".

3.3 Explain what is meant by a "Receiver Operating Characteristic Curve (ROC)", and the "Area under the ROC curve" (AUC).

3.4 Explain the difficulties with assessing unsupervised machine learning models.

4. Be able to evaluate classification models.

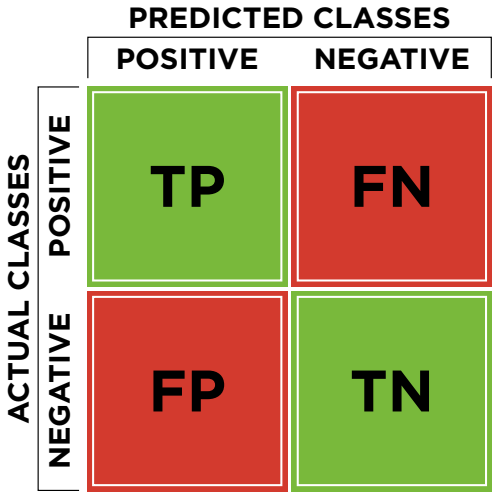
4.1 Calculate the classification metrics from a confusion matrix.

4.2 Interpret a ROC curve and AUC and make reasoned conclusions.

5. Understand the issues of bias and variance in models.

5.1 Explain what is meant by "overfitting" and "underfitting".

5.2 Compare and contrast methods to prevent overfitting: cross validation; removing features; bagging; boosting; early stopping.



Unit 11 – Linear Regression

3 Credits | 30 hours

Unit aims

This unit introduces simple linear regression models that are critical to the ability to predict the value of one continuous variable based on the value of another. Students will be able to estimate the line of best-fit by calculating the regression parameters and understand the accuracy of the model. The unit also introduces multiple and polynomial regression models to examine relationships between multiple variables.

Learning Outcomes and Assessment Criteria

1. Understand the basic theory of linear regression.

1.1 Explain what is meant by simple, multiple, and polynomial linear regression.
1.2 Analyse the assumptions of linear regression.
1.3 Explain the Ordinary Least Squares method for estimating the parameters in simple linear regression.
1.4 Explain the use-cases for linear regression models.
1.5 Evaluate the benefits and limitations of regression models.

2. Understand regression metrics and how to evaluate a regression model.

2.1 Explain the following regression metrics:
- The Total Sum of Squares (TSS).
- The Residual Sum of Squares (RSS).
- The Explained Sum of Squares (ESS).
- The Mean Squared Error (MSE).
- The Root Mean Square Error (RMSE).
- The coefficient of determination (R^2).
- The Adjusted R^2.

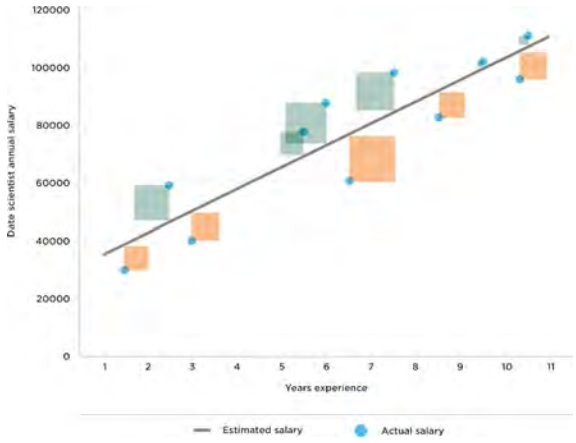
2.2 Explain how to interpret each of the regression metrics listed in 2.1.

3. Be able to perform regression calculations and analysis.

3.1 Calculate the intercept and slope coefficient in simple linear regression.
3.2 Calculate the regression metrics in a linear regression model.
3.3 Interpret the calculated metrics.

4. Be able to create linear regression models.

4.1 Use Python to build accurate simple linear regression and multiple linear regression models for given datasets.
4.2 Use Python to evaluate the accuracy of the models built in 4.1 and analyse the results.



Unit 12 – Logistic Regression

3 Credits | 30 hours

Unit aims

This unit introduces logistic regression and its application as a classification algorithm. The unit explores the basics of binary logistic regression via the logistic function, the Odds ratio, and the Logit function. The unit also explains the differences between linear and logistic regression, teaching the student when it is relevant to choose logistic regression over linear regression.

Learning Outcomes and Assessment Criteria

1. Understand the basic theory of logistic regression.

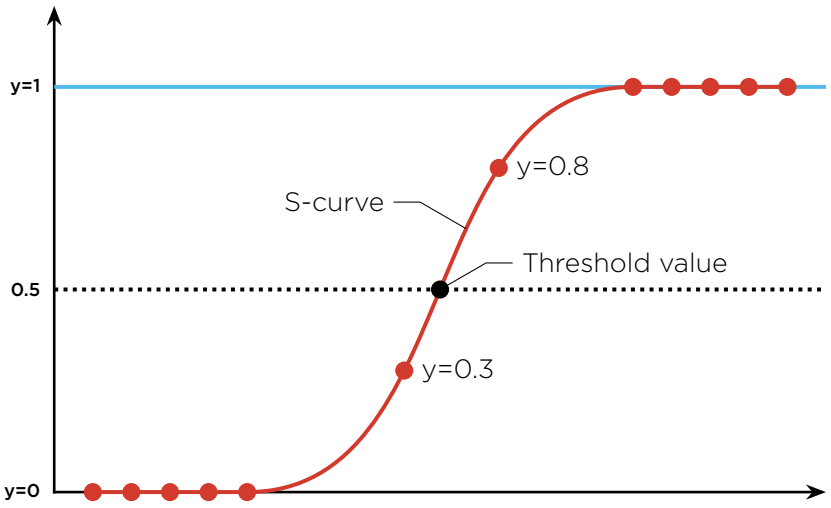
1.1 Explain what is meant by binary logistic regression.
1.2 Explain the difference between linear and logistic regression.
1.3 Analyse the assumptions for logistic regression.
1.4 Define the Logistic function, the Odds ratio, and the Logit function.
1.5 State basic characteristics and properties of the Logistic function, the Odds ratio, and Logit function.
1.6 Explain how to interpret the Odds ratio.
1.7 Evaluate the benefits and limitations of logistic regression.
1.8 Explain how Logistic regression can be applied to multiple-class problems.

2. Be able to perform logistic regression calculations.

2.1 Calculate the probability values of inputs belonging to classes using the Logistic function.
2.2 Calculate the Odds-ratio.
2.3 Calculate relevant classification evaluation metrics for logistic regression model outputs.

3. Be able to create logistic regression models.

3.1 Use Python to build a logistic regression model for datasets.
3.2 Use Python to evaluate the accuracy of the model built in 3.1.
3.3 Analyse the results of the model built in 3.1.



Unit 13 – Decision Trees

3 Credits | 30 hours

Unit aims

This unit introduces the basic theory and application of decision trees. The unit explains how basic classification trees using the standard ID3 decision-tree construction algorithm are built and how nodes are split based on information theory concepts such has Entropy and Information Gain. The student will also build and evaluate decision tree models in Python.

Learning Outcomes and Assessment Criteria

1. Understand what a decision tree is and how it is used.

1.1 Explain the purpose and use-cases of three different decision tree models.
1.2 Discuss the advantages and disadvantages of decision trees.

2. Understand how to construct a decision tree.

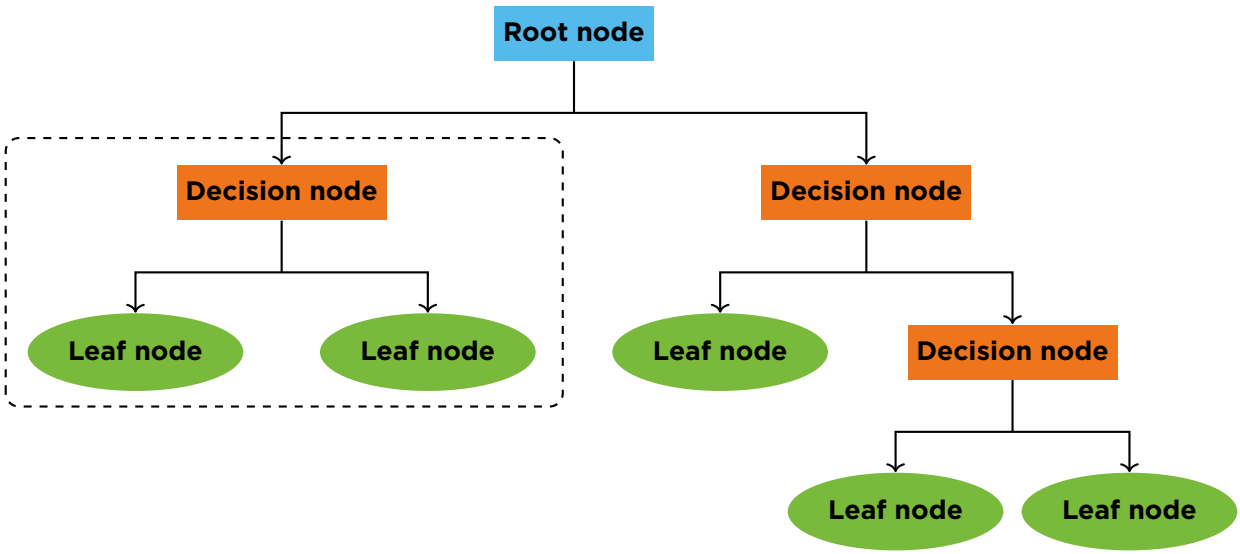
2.1 Explain what is meant by splitting and pruning a decision tree.
2.2 Explain how to calculate and interpret Entropy and Information Gain.
2.3 Explain the key steps in the ID3 (Iterative Dichotomiser) algorithm.
2.4 Discuss improvements and extensions to the ID3 algorithm.

3. Be able to perform calculations using decision tree metrics.

3.1 Calculate Entropy values for a dataset.
3.2 Calculate Information Gain values for a dataset.
3.3 Create visualisations of the Entropy function.

4. Be able to build a decision tree model.

4.1 Use Python to build a decision tree model for a given dataset.
4.2 Use Python to create visualisations for a decision tree.



Unit 14 – k-means clustering

3 Credits | 30 hours

Unit aims

This unit introduces an unsupervised machine learning algorithm: k-means clustering. The unit aims to provide students with the intuition behind k-means clustering and how to find the optimal number of clusters. Finally, the student will also build and evaluate k-means methods in Python, and will learn how to visualise the clusters.

Learning Outcomes and Assessment Criteria

1. Understand the basic theory of k-means clustering.

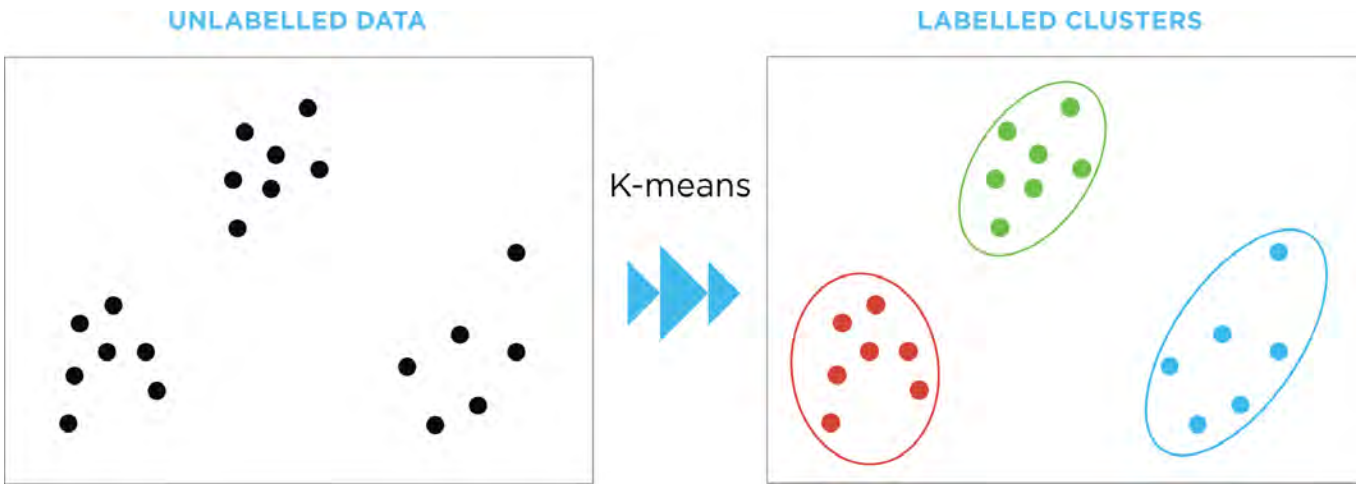
1.1 Explain what is meant by k-means clustering, explaining the terms "cluster" and "centroid".
1.2 Discuss the steps in the k-means clustering algorithm.
1.3 Explain how to determine the optimal number of clusters "k" by using the elbow method.
1.4 Explain how to interpret the Sum of Squared Error (SSE) and The Within-Cluster-Sum of Squared Errors (WSS).
1.5 Discuss the limitations of the elbow method.
1.6 Discuss three types of use-cases k-means clustering can be applied to.
1.7 Evaluate the benefits and limitations of k-means clustering.

2. Understand how to evaluate k-means clusters.

2.1 Explain how to interpret Inertia and Silhouette score.

3. Be able to create and evaluate a k-means model.

3.1 Use Python to build a k-means model.
3.2 Use Python to create visualisations of the clusters generated by the k-means clustering algorithm.
3.3 Use Python to evaluate a k-means model.



Unit 15 – Introduction to Graphs

6 Credits | 60 hours

Unit aims

This unit aims to provide students with an introduction into another emerging area of data science – graphs and graph data science. This unit provides a gentle introduction to the field of graph theory which underpins all modern graph databases and graph analytics. The unit also covers the graph ecosystem, introducing Knowledge Graphs, Labelled Property Graphs and RDF graphs. The unit introduces graph algorithms which are used to model, store, retrieve and analyse graph-structured data.

Learning Outcomes and Assessment Criteria

1. Understand basic types of graphs and their properties.

1.1 Explain what is meant by a "graph", "a vertex", "a node" and an "edge".

1.2 Explain the Bridges of Konigsberg problem and its solution.

1.3 Define the following, providing examples:
- Connected and unconnected graphs.
- Weighted and unweighted graphs.
- Directed and undirected graphs.
- Acyclic and cyclic graphs.
- Monopartite, Bipartite and k-partite graphs.
- Directed Acyclic Graph (DAG).

2. Understand the core types of graph data models.

2.1 Explain what is meant by a "Knowledge graph".

2.2 Explain what is meant by an ontology.

2.3 Explain what is meant by a "Labelled Property Graph" (LPG).

2.4 Explain what is meant by a "Resource Description Framework" (RDF) graph.

3. Understand the graph ecosystem.

3.1 Outline the graph ecosystem from graph databases, graph languages, to graph visualisation tools.

3.2 Compare and contrast LPG databases, RDF databases and relational databases.

3.3 Discuss the use-cases and applications for graph databases.

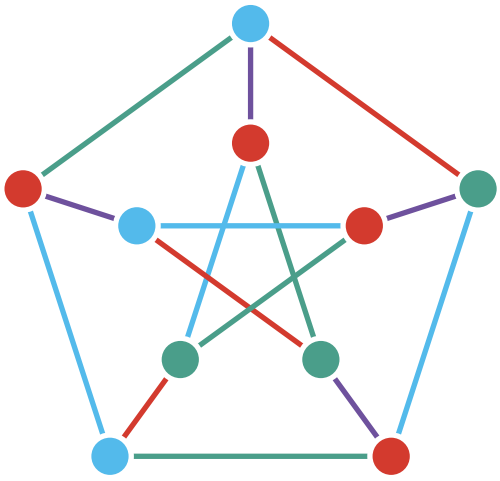
3.4 Discuss Python graph libraries and their features.

4. Understand the types of graph data science and graph algorithms.

4.1 Explain what is meant by "graph data science".

4.2 Discuss the main types of "graph algorithms": search and pathfinding, centrality, and community detection.

4.3 Discuss three types of problems and use-cases that can be tackled by graph data science.



DELTA
ANALYTICS TRAINING

Delta Analytics Training Ltd
Birch House High Street Welford-on-Avon CV37 8EA

01789 750437 support@delta-training.co.uk

Comp 13344598 VAT 420 0438 52