INTERNATIONAL QUALIFICATIONS

AND ASSESSMENT CENTRE (IQAC)

«Approved»

2025 y. «__16__»

Programme	**LEVEL 5 EXTENDED DIPLOMA IN DATA SCIENCE**		
Unit Number/ Unit Title	UNIT 8 DATA MINING AND BIG DATA ANALYTICS		
Cohort Code:	L05DMBD-U8		
Unit Level	Level 5		
Total GLH	Total qualification time 200/ Total Guided learning hours 90/ Self-guided learning hours 110		
Credits	20 CATS/ 10 ECTS		
Lecturer			
Start Date		**End Date**	

Unit Aims	This module aims to introduce students to the concepts and techniques of data mining and analytics for extracting knowledge from large datasets. Students will explore key methodologies in data mining, including clustering, classification, and association rule mining. The module will cover practical applications of these techniques in various domains, enabling students to uncover patterns and insights from large datasets. By the end, they will be equipped to apply data mining tools effectively, enhancing their ability to make data-driven decisions.
Differentiation Strategies *(e.g. planned activities or support for individual learners according to their needs)*	The total number of students to be in the lesson is approximately 20. This is a multicultural group of students predominantly between the ages of 24 – 45, with numerous ethnic, gender, and creed background. These are UK academic level 5 students; hence it is assumed that they have practical, theoretical, or technological knowledge and understanding of a subject or field of work to find ways forward in broadly defined, complex contexts. These students must be able to generate information, evaluate, synthesise the use information from a variety of sources. Various approaches to addressing the various identified students needs will be adopted throughout the lesson. Such will include:- 1. Progressive tasks

	2. Digital resources 3. Verbal support 4. Variable outcomes 5. Collaborative learning 6. Ongoing assessment 7. Flexible-pace learning
Equality & Diversity	Variety of teaching techniques will be employed to ensure that the needs of each individual learner are met.
Safeguarding & Prevent	Safeguarding policies and the Prevent duty are strictly observed to ensure the safety, well-being, and inclusivity of all students and staff.
Health & Safety	SIRM H&S policies will be maintained.
Learning Resources	**Teaching and Learning Materials** • Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann, • Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media. • Zikopoulos, P., Eaton, C., Deutsch, T., & Lapis, G. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media

Learning Outcome		Assessment Criteria
LO1.	**1. Understand data mining concepts and methods.**	1.1 Define data mining and its applications in data science. 1.2 Explain different data mining techniques and algorithms.
LO2.	**2. Develop skills in analyzing big data sets.**	2.1 Apply data preprocessing techniques to prepare data for analysis. 2.2 Implement data mining algorithms to discover patterns and trends in big data.
LO3.	**3. Apply data analytics to real-world scenarios.**	3.1 Evaluate the effectiveness of data mining models. 3.2 Interpret findings and make data-driven decisions based on data analytics results.

No	Learning Outcome / Topic	Learning and Teaching Activities	Which assessment criteria does the session relate to?	Day/month/year/ signature
1.	**Introduction to Data Mining**	**Introduction to Data Mining** Definition, evolution, and applications (e.g., retail, healthcare)	LO1: Data Mining Fundamentals	
2.	**Data Mining Process (CRISP-DM)**	**Data Mining Process (CRISP-DM)** Business understanding → Deployment	LO1: Data Mining Fundamentals	
3.	**Association Rule Mining**	**Association Rule Mining** Apriori algorithm, market basket analysis	LO1: Data Mining Fundamentals	
4.	**Classification Techniques**	**Classification Techniques** Decision trees, Naïve Bayes, k-NN	LO1: Data Mining Fundamentals	
5.	**Clustering Methods**	**Clustering Methods** K-means, DBSCAN, hierarchical clustering	LO1: Data Mining Fundamentals	
6.	**Big Data Characteristics (4Vs)**	**Big Data Characteristics (4Vs)** Volume, velocity, variety, veracity	LO2: Big Data Technologies & Preprocessing	
7.	**Hadoop Ecosystem**	**Hadoop Ecosystem** HDFS, MapReduce, YARN	LO2: Big Data Technologies & Preprocessing	
8.	Half-Term Exam	- Review of LO1 topics - Practice questions and mock assessment - **Half-term assessment** based on LO1 (theory)	LO1 LO2	
9.	**Spark for Big Data Analytics**	**Spark for Big Data Analytics** RDDs, DataFrames, Spark SQL	LO2: Big Data Technologies & Preprocessing	

10.	**Data Cleaning & Transformation**	**Data Cleaning & Transformation** Handling missing values, outliers, normalization	LO2: Big Data Technologies & Preprocessing	
11.	**Feature Engineering**	**Feature Engineering** Dimensionality reduction (PCA), feature selection	LO2: Big Data Technologies & Preprocessing	
12.	**Text Mining & NLP**	**Text Mining & NLP** TF-IDF, topic modeling (LDA), sentiment analysis	LO3: Advanced Data Mining Techniques	
13.	**Time Series Mining**	**Time Series Mining** ARIMA, Prophet, anomaly detection	LO3: Advanced Data Mining Techniques	
14.	Final Exam Preparation & Review	- Comprehensive review of all learning outcomes - Practice questions and revision of key topics		
15.	Final Exam	- **Final-term assessment** covering all learning outcomes (theory and practical elements)		
16.	Feedback & Reflection	- Review of final exam - Individual feedback on performance - Reflective discussion on key learning points		
17.	**Recommender Systems**	**Recommender Systems** Collaborative filtering, content-based filtering	LO3: Advanced Data Mining Techniques	
18.	**Deep Learning for Data Mining**	**Deep Learning for Data Mining** Autoencoders, CNNs for pattern recognition	LO3: Advanced Data Mining Techniques	
19.	**Model Evaluation**	**Model Evaluation** Precision-recall, lift charts, cross-validation	LO3: Advanced Data Mining Techniques	
20.	**NoSQL Databases**	**NoSQL Databases** MongoDB, Cassandra, Neo4j	LO4: Big Data Analytics & Tools	
21.	**Streaming Data Analytics**	**Streaming Data Analytics** Kafka, Spark Streaming, Flink	LO4: Big Data Analytics & Tools	

22.	**Cloud-Based Big Data Solutions**	**Cloud-Based Big Data Solutions** AWS EMR, Google BigQuery, Azure HDInsight	LO4: Big Data Analytics & Tools	
23.	Half-Term Exam	**Project** End-to-end data mining pipeline on a large dataset		
24.	**Graph Mining**	**Graph Mining** PageRank, community detection	LO4: Big Data Analytics & Tools	
25.	**Ethical Considerations**	**Ethical Considerations** Privacy, bias, and GDPR compliance	LO4: Big Data Analytics & Tools	
26.	**Fraud Detection Case Study**	**Fraud Detection Case Study** Anomaly detection in financial transactions	LO5: Real-World Applications & Capstone	
27.	**Customer Churn Prediction**	**Customer Churn Prediction** Telecom industry case study	LO5: Real-World Applications & Capstone	
28.	**Social Network Analysis** and **Future Trends**	**Social Network Analysis** Influencer detection, network graphs **Future Trends** AI-driven automation, federated learning	LO5: Real-World Applications & Capstone	
29.	Final Exam Preparation & Review	LO1, LO2, LO3, LO4	LO1, LO2, LO3, LO4	
30.	Final Exam		LO1, LO2, LO3, LO4	