DEREE COLLEGE SYLLABUS FOR:

ITC 3133 DATA MINING AND BIG DATA **3/0/3**
(Previously: ITC 3233 DATA MINING AND BIG DATA) **UK LEVEL: 5**
(Updated Fall 2023) **UK CREDITS: 15**

PREREQUISITES:	ITC 1070 Information Technology Fundamentals ***or*** ITC 2088 Introduction to Programming
COREQUISITES:	None
CATALOG DESCRIPTION:	Data and feature selection, cleaning, extracting patterns from data, evaluation, big data, tools, applications.
RATIONALE:	The course explores the era of big data and the need to handle the exponentially increasing volumes of data that organizations collect. Students use data mining techniques to navigate through chaotic, heterogeneous, unstructured and noisy data, in order to make inferences. As a result, they develop the necessary skills to proceed, through the use of appropriate tools, with a variety of real-world problems that involve big data, including decision making, marketing, fraud detection, and medicine.
LEARNING OUTCOMES:	As a result of taking this course, the student should be able to: 1. Apply data mining techniques for analysing data and deriving new knowledge. 2. Assess the quality of the inferred information by using a variety of evaluation methods. 3. Combine the appropriate data mining techniques with respect to scalability, to discover information nuggets that are appropriate for a specific problem in a particular domain. 4. Defend the outcomes, in terms of performance, interpretability and visualisation
METHOD OF TEACHING AND LEARNING:	In congruence with the teaching and learning strategy of the college, the following tools are used: • Lectures, class discussions, laboratory practical sessions. • Office hours: Students are encouraged to make full use of the office hours of their instructor, where they can ask questions and go over lecture material. • Use of the Blackboard Learning platform, where instructors post lecture notes, assignment instructions, timely announcements, as well as additional resources.
ASSESSMENT:	Summative:

1st assessment: Coursework short problems	**30%**
2nd assessment: Portfolio of student work and oral assessment.	**10%**
Final assessment: Group Project Programming project to address big data and/or data mining problems.	**60%**

	Formative: Homework, In class quizzes or lab exercises — **0%** The formative assessments aim to prepare students for the summative assessments and expose them to teamwork. The 1st summative assessment tests LOs 1, 2. The 2nd summative assessment tests LOs 1, 2, 3, 4. The final summative assessment tests LOs 1, 2, 3, 4. *The final grade for this module will be determined by averaging all summative assessment grades, based on predetermined weights for each assessment. If students pass the **final summative assessment,** which tests all Learning Outcomes for this module, and the average grade for the module is 40 or above, students are not required to resit any failed assessments.*
INDICATIVE READING:	**REQUIRED MATERIAL:** 1. Witten, I. & Frank, E. (2005), *Data Mining Practical Machine Learning Tools and Techniques*, Elsevier, San Francisco 2. Instructor's notes. **RECOMMENDED READING:** 1. Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Addison Wesley. 2. Hand D., Mannila H., Smyth P., (2001), *Principles of Data Mining*, MIT Press. 3. Zaki, M. J., Meira W.,(2014). *Data Mining and Analysis*, Cambridge University Press.
INDICATIVE MATERIAL: *(e.g. audiovisual, digital material, etc.)*	**REQUIRED MATERIAL:** N/A **RECOMMENDED MATERIAL:** N/A
COMMUNICATION REQUIREMENTS:	Daily access to the course's site on the College's Blackboard CMS. Use of word processing and/or presentation graphics software for documentation of assignments.
SOFTWARE REQUIREMENTS:	Python and related libraries: Scikit-learn, numpy, scipy, matplotlib Apache Flink WEKA
WWW RESOURCES:	• http://www.kdnuggets.com/ • https://www.autonlab.org/resources/tutorials • http://archive.ics.uci.edu/ml/ • http://www.sciencemag.org/site/feature/data/compsci/machine_learning.xhtml
INDICATIVE CONTENT:	1. Introduction to data mining 2. Input 3. Output 4. Classification 5. Validation models 6. Transformations and data pre-processing

	7. Predicting real-valued outputs 8. Clustering 9. Tools for data mining 10. Visualisation in python 11. Big Data and Streaming 12. Further topics in Data mining