

DEREE COLLEGE SYLLABUS FOR:							
ITC 3233 DATA MINING AND BIG DATA (Updated Fall 2021)	3/0/3 UK LEVEL: 5 UK CREDITS: 15						
PREREQUISITES:	ITC 1070 Information Technology Fundamentals <i>or</i> ITC 2088 Introduction to Programming <i>or</i> <i>equivalent</i> MA 2010 Statistics I <i>or</i> MA 2021 Applied Statistics for Business <i>or</i> MA 2025 Applied Statistics for Science						
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: <ol style="list-style-type: none"> 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: <ul style="list-style-type: none"> • 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:	<table border="1"> <tr> <td colspan="2">Summative:</td> </tr> <tr> <td>1st assessment: Coursework short problems</td> <td style="text-align: center;">30%</td> </tr> <tr> <td>2nd assessment: Portfolio of student work and oral assessment.</td> <td style="text-align: center;">10%</td> </tr> </table>	Summative:		1 st assessment: Coursework short problems	30%	2 nd assessment: Portfolio of student work and oral assessment.	10%
Summative:							
1 st assessment: Coursework short problems	30%						
2 nd assessment: Portfolio of student work and oral assessment.	10%						

	<table border="1" data-bbox="643 107 1386 210"> <tr> <td data-bbox="643 107 1292 210">Final assessment: Group Project Programming project to address big data and/or data mining problems.</td> <td data-bbox="1292 107 1386 210">60%</td> </tr> </table> <p data-bbox="643 247 764 275">Formative:</p> <table border="1" data-bbox="643 275 1386 310"> <tr> <td data-bbox="643 275 1208 310">Homework, In class quizzes or lab exercises</td> <td data-bbox="1208 275 1386 310">0%</td> </tr> </table> <p data-bbox="643 365 1448 426">The formative assessments aim to prepare students for the summative assessments and expose them to teamwork.</p> <p data-bbox="643 430 1141 457">The 1st summative assessment tests LOs 1, 2.</p> <p data-bbox="643 462 1200 489">The 2nd summative assessment tests LOs 1, 2, 3, 4.</p> <p data-bbox="643 493 1216 520">The final summative assessment tests LOs 1, 2, 3, 4.</p> <p data-bbox="643 562 1448 758"><i>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.</i></p>	Final assessment: Group Project Programming project to address big data and/or data mining problems.	60%	Homework, In class quizzes or lab exercises	0%
Final assessment: Group Project Programming project to address big data and/or data mining problems.	60%				
Homework, In class quizzes or lab exercises	0%				
INDICATIVE READING:	<p data-bbox="643 800 894 827">REQUIRED MATERIAL:</p> <ol data-bbox="662 831 1377 926" style="list-style-type: none"> 1. Witten, I. & Frank, E. (2005), <i>Data Mining Practical Machine Learning Tools and Techniques</i>, Elsevier, San Francisco 2. Instructor's notes. <p data-bbox="643 961 951 989">RECOMMENDED READING:</p> <ol data-bbox="662 993 1448 1192" style="list-style-type: none"> 1. Tan, P., Steinbach, M., & Kumar, V. (2006). <i>Introduction to data mining</i>. Boston: Pearson Addison Wesley. 2. Hand D., Mannila H., Smyth P., (2001), <i>Principles of Data Mining</i>, MIT Press. 3. Zaki, M. J., Meira W.,(2014). <i>Data Mining and Analysis</i>, Cambridge University Press. 				
INDICATIVE MATERIAL: (e.g. audiovisual, digital material, etc.)	<p data-bbox="643 1236 943 1264">REQUIRED MATERIAL: N/A</p> <p data-bbox="643 1302 1016 1329">RECOMMENDED MATERIAL: N/A</p>				
COMMUNICATION REQUIREMENTS:	<p data-bbox="643 1367 1448 1461">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.</p>				
SOFTWARE REQUIREMENTS:	<p data-bbox="643 1497 1377 1591">Python and related libraries: Scikit-learn, numpy, scipy, matplotlib Apache Flink WEKA</p>				
WWW RESOURCES:	<ul data-bbox="654 1629 1442 1797" style="list-style-type: none"> • 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:	<ol data-bbox="654 1829 1008 1923" style="list-style-type: none"> 1. Introduction to data mining 2. Input 3. Output 				

- | | |
|--|---|
| | <ol style="list-style-type: none">4. Classification5. Validation models6. Transformations and data pre-processing7. Predicting real-valued outputs8. Clustering9. Tools for data mining10. Visualisation in python11. Big Data and Streaming12. Further topics in Data mining |
|--|---|