COLLEGE OF INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE SYLLABUS/SPECIFICATION

CODE & TITLE: **ITCS 431 – Introduction to Data Analytics**

WEIGHT: **(2 - 2 - 3)**

PREREQUISITE: **ITCS 323**

NQF Level Allocated: 8

NQF Notional Hours / Credits: 120 notional hours/ 12 NQF credit

DESCRIPTION:

This course will introduce students to data analytics and equip them with some of its basic principles and tools. Students will learn concepts, techniques, and tools they need to deal with various facets of data analytics. Topics that will be covered include data formats, loading, and cleaning; data storage in relational and non-relational stores; data governance, data analysis using supervised and unsupervised learning using up to date standard tools. In addition, this course would enable students to identify, locate, analyze, and report on business data sources both qualitatively and quantitatively.

OBJECTIVES:

1. To critically analyze and interpret data using appropriate tools to support the decision-making process.
2. To use appropriate models of analysis, assess the quality of input, derive insight from results, and investigate potential issues.
3. To interpret data findings effectively to any audience, orally, visually and in written formats.

SEMESTER: **ACADEMIC YEAR:**

INSTRUCTOR:

OFFICE TEL:

INTENDED LEARNING OUTCOMES (ILOS)

Upon successful completion of the course, students should be able to:

A. Knowledge and Understanding		**NQF Descriptor/Level**
A1	Concepts and Theories: *Demonstrate critical knowledge and understanding of concepts of data analytics.*	Knowledge: theoretical understanding [Level 8]
A2	Contemporary Trends, Problems and Research: *Demonstrate detailed knowledge of the current data* analytics issues and research.	Knowledge: theoretical understanding [Level 8]
A3	Professional Responsibility: NA	

B. Subject-Specific Skills		**NQF Descriptor/Level**
B1	Problem Solving: *Deal with advanced and complex* real data analytics problems and solve them.	Knowledge: Practical Application [Level 8]
B2	Modeling and Design: *Apply standard research to design data* analytics models that meet user requirements.	Knowledge: Practical Application [Level 8]
B3	Application of Methods and Tools: *Apply specialist skills and* advanced data analytics tools to develop data analytics models.	Knowledge: Practical Application [Level 8]

C. Thinking Skills		**NQF Descriptor/Level**
C1	Analytic: *Critically analyze* the performance of the developed data analytics models.	Generic Problem Solving & Analytical skills [Level 8]
C2	Synthetic: *Identify and implement components* of data analytics tools into one complete data analytics model.	Generic Problem Solving & Analytical skills [Level 8]
C3	Creative Thinking and innovation: Demonstrate creativity for solving problems related to data analytics	Generic Problem Solving & Analytical skills [Level 8]

D. General and Transferable Skills (Other Skills Relevant to Employability and Personal		**NQF Descriptor/Level**
D1	Communication: Use special skills to convey complex information and ideas in appropriate oral and written forms.	Communication, ICT and Numeracy Skills [Level 8]
D2	Teamwork and Leadership: Operate Specialist level and work effectively as a member/leader of a team who may plan, design, and implement a data analytics model.	Competence: Autonomy, Responsibility and Context [Level 8]
D3	Organizational and Developmental Skills: Demonstrate ability to organize ideas and effectively allocate time in given assignments.	Competence: Autonomy, Responsibility and Context [Level 8]

D4	Ethical and Social Responsibility: NA	

Course Structure (Outline)						
Week	**Hours**		**ILOs**	**Unit/Module or Topic Title**	**Teaching Method**	**Assessment Method**
	Lec.	**Lab**				
1	2	2	A1,B3	• The Data Science Road Map o Find the problem o Understand the data o Exploratory Analysis o Extract Features	Lecture/ Lab Demonstration	
2	2	2	A1	• Data Munging o String Manipulation o Regular Expressions o Data Cleaning	Lecture	Oral Participation *
3	2	2	A1	• Visualizations and Simple Metrics o Exploratory metrics o Visualization techniques (histograms, bar charts, means, standard deviations, medians, and quantiles)	Lecture	Oral Participation*
4	2	2	A1	• Machine Learning Overview o Background about machine learning	Lecture	Oral Participation*
5	2	2	A1	• Machine Learning Classification o What Is a Classifier? o Binary versus Multiclass o Specific Classifiers o Evaluating Classifiers	Lecture	Oral Participation*
6	2	2	A1, B3	• Unsupervised Learning: Clustering and Dimensionality Reduction	Lecture/ In-Lab Supervised Work	Oral Participation*

				○ Principal Component Analysis and Factor Analysis ○ Factor Analysis ○ Limitations of PCA ○ Clustering		
7	2	2	A1	• Regression ○ Least Squares ○ Fitting Nonlinear Curves ○ Correlation of Residuals ○ Linear Regression ○ LASSO Regression and Feature Selection	Lecture	Oral Participation*
8	2	2	A1, B3	• Data Encodings and File Formats ○ CSV Files ○ JSON Files ○ XML Files ○ HTML Files ○ Tar Files ○ GZip Files	Lecture/ In-Lab Supervised Work	In-Lab Exercices*/ Oral Participation*
9	2	2	A1, B1, B2, B3	• Big Data ○ What Is Big Data? ○ Hadoop: The File System and the Processor ○ Using HDFS ○ Spark Overview ○ Spark Operations ○ Configuring Spark ○ Spark Tips	Lecture/ In-Lab Supervised Work	In-Lab Exercises*/
10	2	2	A1, B1, B2, B3,	• Databases	Lecture/ In-Lab Supervised Work	In-Lab Exercises*/ Oral Participation*/

			C1	o Relational Databases and MySQL o Key-Value Stores o Wide Column Stores o Document Stores		Major Test
11	2	2	A1, A2, B1, B2, B3, D1, D3	• Natural Language Processing o Tokenization o Central Concept: Bag-of-Words o Word Weighting: TF-IDF o n-Grams o Stop Words o Lemmatization and Stemming o Synonyms	Lecture/ In-Lab Supervised Work/ Independent Learning	In-Lab Exercices*/ Oral Participation*/ Assignement 1
12	2	2	A1, B1, B2, B3, C1, C2, C3	• Time Series Analysis o Time Series versus Time-Stamped Events o Resampling an Interpolation o Smoothing Signals o Logarithms and Other Transformations o Trends and Periodicity o Windowing	Lecture/ In-Lab Supervised Work	Lab Project1
13	2	2	A1, B1, B2, B3, C1, C2, C3	• Probability o The Uniform Distribution and Pseudorandom Numbers o Nondiscrete, Noncontinuous Random Variables o Binomial Distribution o Poisson Distribution	Lecture/ In-Lab Supervised Work	Lab Project2 Oral Participation*

				○ Normal Distribution		
14	2	2	A1, A2, B1, B2, B3, D1, D3	• Statistics ○ Bayesian versus Frequentist ○ Hypothesis Testing ○ Multiple Hypothesis Testing ○ Hypothesis Testing: t-Test ○ Bayesian Statistics	Lecture/ In-Lab Supervised Work/ Independent Learning	Assignment 2 (Literature Review)
15	2	2	A2, B1, B2, B3, C1, C2, C3, D1, D2, D3	Student Projects	Project Supervision	Evaluation of Project Presentations and Reports
16	2	-	A1, B1, B2, C1	All Topics		Final Exam

TEACHING MATERIALS:

TEXTBOOK(S):	The Data Science Handbook1st Edition , Field Cady (2017), Wiley
HANDOUT(S):	PowerPoint slides available on Moodle i.e. http://www.ahlia.edu.bh/moodle
REFERENCE(S):	1- Predictive Analytics For Dummies, 2nd Edition, Anasse Bari (2016), For Dummies 2- Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Eric Siegel (2016), Wiley; 2 edition 3- Data Analytics for Beginners: Basic Guide to Master Data Analytics, Paul Kinley (2016), Create Space Independent Publishing Platform.

ASSESSMENTS:

Method of Assessment	Description	Learning Outcomes	Weighting
Oral Participation	Students will be questioned orally to demonstrate their understanding and knowledge of the topics covered during class lectures and lab sessions.	A1	Formative
In-Lab Exercises	Each of the lab exercises consists of a set of practical tasks to be implemented by students individually in class as shown in the above weekly structure. Students work will be observed and directly during the lab sessions.	B3	Formative
Assignments	Two assignments for Literature review. Average will be taken.	A2, D1, D3	15%
Lab Projects	Students will be asked to use and apply appropriate development tools to develop data analytics models. Students has to code, test and deploy interactive data analytics models.	B1, B2, B3, C1, C2, C3	10%
Major Test	The test will be an in-class 90 minutes exam that will consist of short-answer, essay, and problem solving questions and cover the topics studied in the first 8 weeks.	A1, B1, B2, C1	20%
Project	Each group of 2-4 students has to develop a data analytics models for solving a real world problem. Each group has to go through all phases of system development cycle, and submit a report and present the work in the class.	A2, B1, B2, B3, C1, C2, C3, D1, D2, D3	15%
Final Exam	The final exam is comprehensive and will be of two hours duration. It will consist of short-answer, analysis, essay, and problem-solving questions.	A1, B1, B2, C1	40%
Overall:			**100 %**

14. Admissions	
Pre-requisites	ITCS 323
Minimum number of students	8
Maximum number of students	20