

COURSE STRUCTURE

Name of Course: INTRODUCTION TO STATISTICS AND DATA ANALYTICS

Course Code: DCS1104

Credit Hours: 4

Prerequisite/co-requisite: None

Summary:

This course exposes students to the concepts of data analytics. It explains how to visualize and summarize data, and how to find natural groups and frequent patterns in a dataset using R language. It also explores clustering, predictive tasks, classification and regression.

Course Learning Outcomes:

Upon completing this course, the students will be able to:

CLO1: Explain the concepts of data analytics. (C2, PLO1)

CLO2: Display the ability to visualize different broad range data sets. (P3, PLO3)

CLO3: Display the ability to execute predictive, clustering, classification, and regression using a modern example. (P3, PLO3)

CLO4: Show the ability to execute and run R Studio. (P3, PLO3)

Course Format:

Total Student Learning Time (SLT) (L = Lecture; T = Tutorial; P = Practical; EL = E-Learning) :					
Learning Hours				Guided & Independent Learning (hr)	Total Student Learning Time (hr)
L	T	P	EL		
28	0	24	14	94	160

Teaching and Delivery Methods/ Teaching Methodology:

Lectures, Tutorial and Practical/Laboratory work delivered in a combination of blended & independent learning

E-Learning provided by INTI makes learning more accessible and convenient for the students. The blended model utilized by INTI is the integration of E-learning via INTI's Learning Management System and the conventional lecturer-led classroom activities. INTI students are required to access to the online learning materials (additional notes, reading materials, online assessments, discussion forums and etc.), so as to acquire a complete learning process. This also promotes self-directed learning in encouraging INTI students to be independent learners.

Syllabus:

	Course Content Outline	CLO*
1 - 2	Introduction to Data What Can We Do With Data, Big Data and Data Science, Big Data Architectures, Small Data, What is Data?, A Short Taxonomy of Data Analytics, Examples of Data Use.	1
3 - 4	Getting Insights from Data: Descriptive Statistics Scale Types, Descriptive Univariate Analysis, Univariate Frequencies, Univariate Data Visualization, Univariate Statistics, Common Univariate Probability Distributions, Descriptive Bivariate Analysis, Two Quantitative Attributes, Two Qualitative Attributes	1,2
7 – 8	Descriptive Multivariate Analysis Multivariate Frequencies, Multivariate Data Visualization, Multivariate Statistics, Location Multivariate Statistics, Dispersion Multivariate Statistics, Infographics and Word Clouds	1,2
9 – 12	Data Quality and Preprocessing Data Quality, Missing Values, Redundant Data, Inconsistent Data, Noisy Data, Outliers, Converting to a Different Scale Type, Converting Nominal to Relative, Converting Ordinal to Relative or Absolute, Converting Relative or Absolute to Ordinal or Nominal, Converting to a Different Scale, Data Transformation	2
13 - 14	Clustering Distance Measures, Differences between Values of Common Attribute Types, Distance Measures for Objects with Quantitative Attributes, Distance Measures for Non-conventional Attributes, Clustering Validation, Clustering Techniques	3,4
15 – 18	Frequent Pattern Mining Frequent Itemsets, Setting the min_sup Threshold, Apriori – a Join-based Method, Eclat, FP-Growth	3,4
19 - 20	Predicting the Unknown Regression, Predictive Performance Estimation, Generalization, Model Validation, Predictive Performance Measures for Regression, Finding the Parameters of the Model, Linear Regression	3,4
21 - 24	Classification Binary Classification, Predictive Performance Measures for Classification	3,4
25 - 28	Classification Distance-based Learning Algorithms, K-nearest Neighbor Algorithms, Case-based Reasoning	3,4
	FINAL EXAMINATION	

Student Evaluation:

Continuous Assessment		Percentage (%)	CLO
1	Test	20	1
2	Lab Report	20	2
3	Assignment	20	3,4
Total		60%	
Final Assessment		Percentage (%)	
Final Examination		40	1

Final exam format:

Duration: 2 hours

Students are required to answer FOUR Structured questions. All questions carry equal marks.

Grading Scale:

A+ (90-100), A (80-89), A- (75-79), B+ (70-74), B (65-69), B- (60-64), C+ (55-59), C (50-54), C- (45-49), D (40-44), F (0-39), Resit Pass, RP (50-100), Resit Fail, RF (0-49)

IMPORTANT NOTE:Students are required to “**PASS**” BOTH continuous and final assessment in order to pass the subject.**Additional Information:** R Studio**Main Reference(s) Supporting Course:**

1. Daniel Bell (2019), Independent Publisher, R Programming, A Step-by-Step Guide for Absolute Beginners. ISBN: 978-1696769648

Additional References:

1. João Mendes Moreira André C. P. L. F. de Carvalho Tomáš Horváth, (2018), A General Introduction to Data Analytics, 1st Edition, John Wiley & Sons, Inc, ISBN: 9781119296249
2. Richard Hurley (2019), Data Science: A Comprehensive Guide to Data Science, Data Analytics, Data Mining, Artificial Intelligence, Machine Learning and Big Data (Kindle Edition)

LABORATORY WORK:

Lab	Practical Work
1	Installing R- Studio
2	Importing data
3	Understanding different types of data
4	Graphical Representation of data
5 – 6	Descriptive statistics on data
7 - 8	Descriptive multivariate analysis
9 – 10	Data pre-processing
11	Classification algorithm I
12	Classification algorithm II