****

Course Title: Foundations of Data Science Using Programming Language

Course Code: IS-621

Resource Person:

Department: **Information Systems**

**HSM Vision**

HSM envisions its success in the sustainable contribution that it will make to the industry, academia and research in public and private sector. HSM will lead by providing professionally competent and ethically conscious human resources engaged in the global and local context to foster socio-economic growth and sustainability for the society. HSM envisages having faculty with high research potential and a deep desire for cutting edge research including collaboration with national and international partners.

**HSM Mission**

Being a research-oriented and student-centric business school, we emphasize research publications in impact journals as well as state-of -the-art learning methodologies.  We will prepare our students to become the future ethical business leaders and the guiding post for the society, while equipping them with the knowledge and skills required by world-class professionals.  We will be the leading choice for organizations seeking highly talented human resource. HSM will foster internationalization with key stakeholders and actively work to exchange best practices with business schools across Pakistan through collaborations, workshops, conferences and other means.

**Program Objectives**

**Course Objectives**

This course will teach how to program in R and how to use R for effective data analysis. The students will learn how to install and configure R necessary for an analytics programming environment and gain basic analytic skills via this high-level analytical language. The course covers fundamental knowledge in R programming. Popular R packages for data science will be introduced as working examples. The format of the course will include lectures by the instructor, computing labs, class discussion, directed reading, and student presentation or project. The exact format will depend on the size of enrolment and student background. By doing data analysis using R tool able to create thousands of virtual experiments based on input data, researchers are able to ‘game out’ and then model the impacts of various changes to processes or chemical compounds, far more quickly and cost-effectively. This is especially valuable during the early stages of R&D. The importance of data analysis is greater when considering the urgency around issues related to climate change.

**Learning Objectives**

* Understanding the general concepts of Data Analysis and its different terminologies.
* Discuss recent applications of Data Analysis.
* Understanding of technical issues related to robotic control
* Understanding different applications used in data mining as well as autonomous navigation.
* Detail understanding on, bioinformatics, and speech recognition text and web data processing.

**Pre-requisites:**

Some basic knowledge of programming, probability and statistics. If in doubt about the prerequisites, please consult with the instructor for permission to take the class.

**Teaching Methodology (List methodologies used –example are given below)**

* Lecture
* Interactive Classes
* Case based teaching
* Class activities
* Applied Projects
* Guest Lectures

**STUDENTS ARE REQUIRED TO READ AND UNDERSTAND ALL ITEMS OUTLINED IN THE PARTICIPANT HANDBOOK**

**Class Policy:-**

* Be On Time

You need to be at class at the assigned time. After 10 minutes past the assigned time, you will be marked absent.

* Mobile Policy

**TURN OFF YOUR MOBILE PHONE!** It is unprofessional to be texting or otherwise.

* Email Policy

**READ YOUR EMAILS!** You are responsible if you miss a deadline because you did not read your email.

Participants should regularly check their university emails accounts regularly and respond accordingly.

* Class Attendance Policy

A minimum of 80% attendance is required for a participant to be eligible to sit in the final examination. Being sick and going to weddings are absences and will not be counted as present. You have the opportunity to use 6 absences out of 30 classes. Participants with less than 80% of attendance in a course will be given grade ‘F’ (Fail) and will not be allowed to take end term exams. International students who will be leaving for visa during semester should not use any days off except for visa trip. Otherwise they could reach short attendance.

* Withdraw Policy

Students may withdraw from a course till the end of the 12th week of the semester. Consequently, grade W will be awarded to the student which shall have no impact on the calculation of the GPA of the student.A Student withdrawing after the 12th week shall be automatically awarded “F” grade which shall count in the GPA.

* Moodle

UMT –LMS (Moodle) is an Open Source Course Management System (CMS), also known as a learning Management System (LMS). Participants should regularly visit the course website on MOODLE Course Management system, and fully benefit from its capabilities. If you are facing any problem using moodle, visit <http://oit.umt.edu.pk/moodle>. For further query send your queries to moodle@umt.edu.pk

* Harassment Policy

Sexual or any other harassment is prohibited and is constituted as punishable offence. Sexual or any other harassment of any participant will not be tolerated. All actions categorized as sexual or any other harassment when done physically or verbally would also be considered as sexual harassment when done using electronic media such as computers, mobiles, internet, emails etc.

* Use of Unfair Means/Honesty Policy

Any participant found using unfair means or assisting another participant during a class test/quiz, assignments or examination would be liable to disciplinary action.

* Plagiarism Policy

All students are required to attach a “Turnitin” report on every assignment, big or small. Any student who attempts to bypass “Turnitin” will receive “F” grade which will count towards the CGPA. The participants submit the plagiarism report to the resource person with every assignment, report, project, thesis etc. If student attempts to cheat “Turnitin”, he/she will receive a second “F” that will count towards the CGPA. There are special rules on plagiarism for final reports etc. all outlined in your handbook.
* Communication of Results

The results of quizzes, midterms and assignments are communicated to the participants during the semester and answer books are returned to them. It is the responsibility of the course instructor to keep the participants informed about his/her progress during the semester. The course instructor will inform a participant at least one week before the final examination related to his or her performance in the course.

**Course Outline**

Course code: Course title: Foundations of Data Analysis in R

|  |  |
| --- | --- |
| Program |  |
| Credit Hours | 3 |
| Duration | 15 Weeks |
| Prerequisites (If any) | N/A |
| Resource PersonName and Email |  |
| Counseling Timing(Room# 1N1 R#7 ) |  |
| Contact no. |  |
| Web Links:-(Face book, Linked In, Google Groups, Other platforms) |  |

**Chairman/Director Program signature………………………………….Date……………………..**

**Dean’s signature………………………… ………………….Date………………………………………….**

**Grade Evaluation Criteria**

Following is the criteria for the distribution of marks to evaluate final grade in a semester.

**Marks Evaluation Marks in percentage**

Quizzes 20%

Assignments and Lab Work 10%

Mid Term 25%

Term Project and Presentation 15%

Final Exam 30%

Total 100%

**Reference Books:**

* R Programming for Data Science, by Roger D. Peng, https://leanpub.com/rprogramming
* Using R for Introductory Statistics, by John Verzani, Chapman & Hall/CRC, 2004, ISBN 1584884509
* Advanced R, by Hadley Wickham, ISBN 9781466586963.

**Software**

R programming language use to implement some learning algorithms during data analysis.

**Course: -** **Foundations of Data Analysis in R Course code: Book:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Topics to be****covered in the course** | **Learning Objective****of this topic** | **Expected Outcomes from Students** | **Teaching Method** | **Assessment Criteria** | **Deadlines and Homework** |
| 1 | Introduction R Nults and Bolts (I) | Basic Concept of R programming Language.  | Students will be able to start with the basic terminology doing Data Analytics using R Language.  | Lecture | Assignment | Within a Week |
| 2 | R Nults and Bolts (II) | Understanding the concept of R Nults and Bolts | Students will be able to setup R Nults and Bolts | Case Discussion | Assignment | Within a Week |
| 3 | Getting Data In and Out of R | Understanding the basic functionality of extracting the data and importing the data in R programming Language | They will be in position to understand the key method of importing and exporting the data.  | Case Discussion + Lecture | Class Activity | Within a Week |
| 4 | Control Structures and Functions | Learn the basics of Control Structures and Functions in order to perform Data Analytics. | After this lecture, students will be able to determine multiple Control Structures and Functions. | Lecture + Lab work | Quiz | Within a Week |
| 5 | Loop Functions | Working on loop functions. Understanding the use of loop functions. | They will be able to work on basic functionality of loop Function. | Case Discussion + Lecture | Class Activity | Within a Week |
| 6 | Data Manipulation (dplyr, reshape2 packages) | Understand the Data Manipulation including dplyr, reshape2 packages.  | Students will understand Data Manipulation | Case Discussion + Lab Work | Quiz | Within two Weeks |
| 7 | String Operations (stringr package) | Providing in depth details on String Operations specifically how to use stringr package. | After this, students will learn how to use String Operations. | Lecture  | Class Activity | Within two Weeks |
| 8 | Mid Term | Mid Term | Mid Term | Mid Term | Mid Term | Mid Term |
| 9 | Packaging, Debugging and Object Oriented Programming | Learn the basics about packages , Debugging and Object Oriented Programming that used in R language | After this lecture, students will be able to understand the basics packages , Debugging and Object Oriented Programming that used in R language | Discussion + Lab Work | Assignment | Within a Week |
| 10 | Data Visualization (ggplot2 package) | How to perform Data Visualization | Now they will be able to Data Visualization | Lecture + Case Study | Class Activity | Within a Week |
| 11 | Clustering | Understanding the Clustering. K-means and EM | After this lecture, students have understood various techniques of performing Clustering  | Lab Work + Case Study | Quiz | Within a Week |
| 12 | Regression and Classification | Understanding Regression as well as different Classification used in Regression. | Their understanding with various Regression as well as different Classification used in Regression will improve. | Lecture + Case Discussion | Assignment | Within two Weeks |
| 13 | Data Analytics Case Study | Evaluating Data Analytics by discussing different real time case studies. | They will be able to understand different perspectives of using Data Analytics. | Lab Work + Case Study | Quiz | Within two Weeks |
| 14 | Project Presentation | Presentation | Presentation | Presentation | Class Activity | Within a Week |
| 15 | Guest Lecture and QA session with Data Analytics |  | How industry focuses on Data Analytics in order to create sustainable environment  | Guest Speakers Presentation | QA session, Discussion  |  |