**Introduction to Data Science-IS-371**

|  |  |
| --- | --- |
| Resource Person: | Ms. Amna Altaf |
| Email:  | Amna.altaf@umt.edu.pk |
| Contact Hours: | Monday: 12:00-1:00Wednesday: 2:00-3:00Thursday: 12:00-1:00 |
| Office Address: | New faculty hall, Cabin no. 3 |
| Programme: | BBA(h) |
| Section: |  |
| Semester: | Fall 2022 |
| Course Pre-requisites: |  |
| Credit Hours: | 03 |
| Course Type: | Mixed: Theory and Practice |
| Venue/Day/Time: |  |
| Course URL (if any): |  |

|  |
| --- |
| **Course Description:** |
| Today, businesses, consumers, and societies leave behind massive amounts of data as a by-product of their activities. Leading-edge companies in every industry are using analytics to replace intuition and guesswork in their decision-making. As a result, managers are collecting and analyzing enormous data sets to discover new patterns and insights and running controlled experiments to test hypotheses.This course prepares students to understand data science and become leaders in their organizations. This course teaches the scientific process of transforming data into insights for making better business decisions. It covers the methodologies, issues, and challenges related to analyzing business data. It will illustrate the processes of data science by allowing students to apply data science algorithms and methodologies to business problems.**Learning Objectives*** Understand basic concepts of statistics and probability and their application in understanding behavior of data.
* Apply basic tools for performing exploratory data analysis and visualization.
* Understand basic predictive modeling and data analysis methods.
* Learn Python for performing different data science steps.
 |

|  |
| --- |
| **Course Teaching Methodology:** |
| Interactive Classes |

|  |
| --- |
| **Programme Educational Objectives (POs):** |
| PO-1 | To develop effective Teamwork and Leadership Skills |
| PO-2 | To inculcate Critical Thinking and effective Decision-Making skills |
| PO-3 | To develop Effective Communication Skills |
| PO-4 | To polish Core Business Knowledge and Competence |
| PO-5 | To expose and inculcate Ethical Behavior and Social Responsibility |
| PO-6 | To provide real-life work experiences. |
| PO-7 | To provide global perspectives. |

|  |
| --- |
| **Programme Learning Outcomes (PLOs):****After completing this degree programme, students shall be able to:** |
|  | **Mapping of PLO’s with PO’s** |
| PLO1 | Work effectively in teams and understand group processes, leadership, conflict, power and culture in organization. | PO1, PO5, PO6 |
| PLO2 | Use analytical and reflective thinking techniques. | PO2, PO4, PO6 |
| PLO3 | Apply appropriate quantitative and qualitative techniques in solving business problems. | PO2, PO3, PO4, PO5, PO6 |
| PLO4 | Draft effective business documents and prepare and deliver effective oral business presentations using the variety of appropriate technologies. | PO1, PO3, PO6 |
| PLO5 | Demonstrate competency in the underlying concepts, theory and tools taught in the core undergraduate curriculum. | PO4, PO5, PO6, PO7 |
| PLO6 | Identify and analyze ethical conflicts and social responsibility issues involving different stakeholders. | PO5, PO6 |
| PLO7 | Understand the dynamics of industry and understand business as an integrated system and apply strategic planning tools. | PO2, PO3, PO6 |
| PLO8 | Identify and analyze relevant global factors that influence decision making in an international business setting. | PO6, PO7 |

|  |
| --- |
| **Course Learning Outcomes (CLOs):****After completing this course, students shall be able to:** |
|  | **Mapping the CLOs with PLOs** |
| CLO-1 | Apply data analysis techniques to solve business problems and make informed decisions. | PLO-3 |
| CLO-2 | Use analytical and reflective thinking techniques to evaluate data and draw meaningful insights. | PLO-2 |
| CLO-3 | Collaborate effectively in teams to analyze and interpret data, and present findings to stakeholders. | PLO-1 |
| CLO-4 | Draft clear and concise business documents, such as reports and presentations, to effectively communicate data-driven insights. | PLO-4 |
| CLO-5 | Demonstrate competency in the core concepts and theories of data science, including data collection, preprocessing, analysis, and visualization. | PLO-5 |
| CLO-6 | CLO-6: Identify and analyze ethical considerations and social responsibility issues in data science practices. | PLO-6 |
| CLO-7 | CLO-7: Understand the impact of global factors on data-driven decision making and evaluate their influence on business strategies. | PLO-8 |
| CLO-8 | Apply different analysis techniques to identify competitive advantages and opportunities for business growth in dynamic market environments. | PLO-7 |

|  |
| --- |
| **Assurance of Learning and Assessment Items:***Specify Assessment Items that will assure student learning through application and achieve objectives of specific PLOs / COs / CLOs* |
| **Assessment Item** | **Application/ Objectives****PLO / CO / CLO** |
| Quizzes | CLO-1, CLO-2, CLO-8, CLO-5, CLO-6 |
| Assignments | CLO-5, CLO-6, CLO-8, CLO-1 |
| Mid-term exam | CLO-1, CLO-2, CLO-3 |
| Final exam | CLO-8, CLO-6, CLO-7 |
| Project | CLO-4, CLO-6, CLO-7, CLO-3 |
| Class Participation | CLO-7, CLO-1 |

|  |
| --- |
| **Assessment Structure and Grading Policy\*:** |
| **Assessment Item** | **Weight (%)** | **Execution Plan** |
| Quizzes | 10 | 04 |
| Assignments | 10 | 04 |
| Mid-term exam | 25 | One-time assessment |
| Final exam | 25 | One-time assessment |
| Project | 25 | One-time assessment |
| Class Participation | 5 |  |
| **Total**  | **100** |  |
| **Notes – Norms and Important Class Policies:** *(such as submission guidelines, academic honesty, make-up policy, code of conduct)*• Be On TimeYou need to be at class at the assigned time. After 10 minutes past the assigned time, you will be marked absent. • Mobile PolicyTURN OFF YOUR MOBILE PHONE! It is unprofessional to be texting or otherwise.• Email PolicyREAD 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 PolicyA 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 PolicyStudents 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 PolicySexual 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 PolicyAny participant found using unfair means or assisting another participant during a class test/quiz, assignments or examination would be liable to disciplinary action. • Plagiarism PolicyAll 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 ResultsThe 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.  |

*\*Rubrics for all assessments (including mid and final exams) will be provided separately to the students.*

|  |
| --- |
| **Weekly Sessions Plan:** |
| **Week** | **Topics / Contents** | **Activity** | **Application/Objectives****PLO / CO / CLO** |
| 1 | **Introduction of Data Science** * What is data science?
* Importance of data science?
* How data science is transforming business?
 | **Lab Work**: Google Collab Tutorial/ Assignment 1 | CLO-5 |
| 2 | **Data Science Process** * What is CRISP DM model?
* How data is collected?
* What are missing values how to deal with them?
 | **Lab Work**: Basics of Python/ Class Discussion | CLO-5, CLO-6 |
| 3 | **Data Science Process** * What is data modeling. What is curse of dimensionality?
* How to deal with outliers?
* How discovered insights help in real world applications?
 | **Lab Work**: Basics of Python/Assignment 2 | CLO-5, CLO-6 |
| 4 | **Basics of Machine Learning** * Difference between Inductive learning and deductive learning
* Balanced data and unbalanced data annotated, unannotated, and semi-annotated data
* Supervised, unsupervised, and semi-supervised machine learning
* Class balanced split and random split
* Training regimes
* Phases of machine learning
 | **Lab Work**: Basics of Python/Quiz 1 | CLO-5, CLO-6 |
| 5 | **Market Basket Analysis** * Real world application of MBA
* Association rule mining using apriori algorithm.
 | **Lab Work**: Pandas Library/Preprocessing | CLO-2, CLO-7, CLO-8 |
| 6 | **Linear Regression*** How to find relationship between dependent and independent variable.
* Equation of linear regression, error estimation.
* Representation of input and output for linear regression.
 | Quiz -2 | CLO-1, CLO-2, CLO-8 |
| 7 | **Multiple Linear Regression Analysis** * How to find relationship between dependent and independent variables?
* Equation of multiple linear regression.
* Error estimation.
* Representation of input and output for multiple linear regression.
 | **Lab Work**: Implementation of multiple linear Regression/Assignment 4 | CLO-1, CLO-2, CLO-8 |
| 8 | **Logistic Regression*** Draw backs of linear regression for binary output values, use of sigmoid function, loss function.
 | **Lab Work**: Implementation of Logistic Regression | CLO-1, CLO-2 |
| 9 | **Mid Term Exam** |  |  |
| 10 | **Exploratory Data Analysis** * Data understanding,
* visualize outliers
* Different types of graphs for data visualization
 | **Lab Work**: Data visualization / Assignment 3 | CLO-1, CLO-2, CLO-8 |
| 11 | **Classification*** Binary, ternary, multiclass, multilabel, Content based features for text, stylometry features for text.
 | **Lab Work**: Feature extraction (content based and stylometry) for text classification | CLO-1, CLO-2 |
| 12 | **Evaluation Measures*** Precision
* Recall
* F1-score
* Accuracy
* Confusion matrix,
* Comparison of classifiers
* Cross validations.
 | Lab work: Confusion matrix/ Quiz 3 | CLO-1, CLO-2 |
| 13 | **KNN Classifier*** Calculation of Euclidian Distance
* Benefits of KNN
* Numerical
 | **Lab Work**: Implementation of KNN  | CLO-1, CLO-2 |
| 14 | **Naïve Bayes Working** * **A**ssumptions of naïve bayes
* Bayes theorem (derivation)
* Dealing with missing values
* Laplace smoothing .
 | Quiz 4/Revision | CLO-1, CLO-2 |
| 15 | Project Presentation & Viva | Class Discussion | CLO-3, CLO-4 |
| 16 | Final Term Examination |  |  |

|  |
| --- |
| **Primary Text Book (s):** |
| Data Science Concept and Practice, 2md Edition |

|  |
| --- |
| **Reference / Supplementary Reading (s):** |
| Machine Learning by Tom M. Mitchell |

|  |
| --- |
| **Useful Online / Web Resources:** |
| * https://lms.umt.edu.pk/
 |